THE UNIVERSITY
of EDINBURGH

Data-driven modelling for demand response from large consumer energy assets

Gautham Krishnadas

THE UNIVERSITY *of* EDINBURGH

Thesis submitted for the degree of

Doctor of Philosophy

School of Engineering

2018

Lay summary

Demand response (DR) refers to the changes in energy consumption/generation at the consumer-side based on signals/notification from the electricity system. DR acts as a resource that can be used when the electricity supply and demand balance is at risk. For example, when a generator fails (referred to as an event), the system operator sends signals to a **large consumer** such as a supermarket building (and many others) to reduce their energy consumption. The consumers respond by lowering/switching-off their *flexible loads* (refrigerators, space heaters), running a *standby generator* or operating a *battery* backup. These *elements* are referred to as **energy assets** and the consumers get financially rewarded for their contribution.

In many electricity systems, supply and demand imbalance issues are on the rise. One of the primary reasons behind this is the increasing volumes of uncertain and intermittent renewable energy generation. Wider adoption of electric vehicles and ad hoc charging could aggravate this in the near future. Since DR is the most cost-effective solution to solving the grid balancing problem, its requirement is higher than ever before. Development of novel techniques to enhance DR activities is essential in this context.

The smart electricity systems of today produce large amounts of data such as generation, consumption, emissions, grid frequency, among others. However, these data are seldom utilised for better decision making in the electricity system operations. The presented thesis proposes and critically analyses the use of data for improving DR which is an important mechanism in the smart electricity system operations. Primarily, **data-driven modelling** is performed for large consumer building load estimation with specific focus on computational efficiency and potential for large-scale deployment. Following this, application of the data-driven model in different DR activities such as capacity scheduling (before an event), reliable operation (during an event) and performance evaluation (after an event) is demonstrated using the frameworks of ongoing DR programs. The thesis provides conclusive evidences on how the data-driven models prove to be more reliable than the conventional models used in these activities. Apart from these, the research also highlights the use of real-time emissions data to encourage large consumers in shifting their energy consumption away from carbon-intensive generation (such as coal based thermal power).

Abstract

Demand response (DR) is one of the integral mechanisms of today's smart grids. It enables consumer energy assets such as flexible loads, standby generators and storage systems to add value to the grid by providing cost-effective flexibility. With increasing renewable generation and impending electric vehicle deployment, there is a critical need for large volumes of reliable and responsive flexibility through DR. This poses a new challenge for the electricity sector.

Smart grid development has resulted in the availability of large amounts of data from different physical segments of the grid such as generation, transmission, distribution and consumption. For instance, smart meter data carrying valuable information is increasingly available from the consumers. Parallel to this, the domain of data analytics and machine learning (ML) is making immense progress. Data-driven modelling based on ML algorithms offers new opportunities to utilise the smart grid data and address the DR challenge.

The thesis demonstrates the use of data-driven models for enhancing DR from large consumers such as commercial and industrial (C&I) buildings. A reliable, computationally efficient, cost-effective and deployable data-driven model is developed for large consumer building load estimation. The selection of data pre-processing and model development methods are guided by these design criteria. Based on this model, DR operational tasks such as capacity scheduling, performance evaluation and reliable operation are demonstrated for consumer energy assets such as flexible loads, standby generators and storage systems. Case studies are designed based on the frameworks of ongoing DR programs in different electricity markets. In these contexts, data-driven modelling shows substantial improvement over the conventional models and promises more automation in DR operations. The thesis also conceptualises an emissions-based DR program based on emissions intensity data and consumer load flexibility to demonstrate the use of smart grid data in encouraging renewable energy consumption.

Going forward, the thesis advocates data-informed thinking for utilising smart grid data towards solving problems faced by the electricity sector.

Acknowledgements

I am deeply grateful to Marie Skłodowska Curie actions and the team behind project ADVANTAGE (Advanced communications and information processing in smart grid systems) without whom this research would have been a distant dream. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 607774.

A major part of the research work was conducted at Flexitricity Limited, a smart grid services company based in Edinburgh and one of the industrial collaborators with ADVANTAGE. I would like to thank Dr. Alastair Martin (Founder, Flexitricity) for helping me maintain a good balance between the academic demands of the doctoral program and the business priorities of the company – ''*working with you and learning from you will be part of my proudest memories*''. The support and warmth extended by my colleagues at Flexitricity have created the most congenial environment for this work – *''thanks for making me feel at home in Edinburgh; and for the many tea rounds that have arguably made me an addict by now!''.*

During the doctoral program, I was affiliated to the Institute for Energy Systems at the University of Edinburgh. I would like to thank Dr. Aristides Kiprakis for helping steer this research in the proper direction through many hours of insightful discussions and necessary feedbacks – *''you have been the friendliest and the most supportive supervisor I could ever wish for''*.

Prof. John Thompson has been a great source of inspiration as a cheerful leader and a strong pillar for ADVANTAGE. Ms. Hazel Cox has provided the most skilful administrative support. To John, Hazel and all the researchers who were part of the team ADVANTAGE – *''thanks for this wonderful opportunity; it was fun collaborating through the training events, science fairs, conferences and project deliverables''*.

As part of a secondment, four memorable months were spent at the Rocky Mountain Institute (RMI) in Boulder-Colorado, an organisation synonymous with sustainable development. Mr. Mark Dyson and other colleagues at RMI were of immense help during this stint. The research work conducted, sitting at the foot of the Rockies, has been included in the thesis.

The valuable comments from Dr. Dimitrios Rovas and Dr. Chris Dent as the examiners have helped improve the thesis presentation and I am very thankful for their precious time. I am also indebted to the following people for their academic guidance at different points in my life: Dr. T.V Ramachandra, Prof. Sjef Cobben, Dr. Rajendra Babu and Dr. Deepa Mangalam.

To my brother, sister, family and friends – *''life would have been so mundane without you all''*.

Aparna – *''any words of acknowledgement would fall short for your love, understanding, patience and company; all I could offer in return were the sermons on smart grids!''*.

To my parents – *''you have been the perennial sources of encouragement and strength throughout this journey, having taught me the value of knowledge quite early in life, for which I am always grateful''*.

Almighty – *''you know it all''*.

Declaration

I declare that this thesis has been composed by myself and the work presented is entirely my own. The thesis has not been submitted, in whole or in part, for any other degree or professional qualification.

August 2018

Gautham Krishnadas

Contents

List of figures

List of tables

List of abbreviations

A

ADVANTAGE	Advanced Communications and Information processing in smart grid systems
ANN	Artificial neural networks

B

BESS	Battery energy storage system

C

C&I	Commercial & Industrial
CAISO	California independent system operator
CART	Classification and regression trees

D

DoD	Depth of discharge
DR	Demand response
DRUPS	Diesel rotary uninterrupted power supply
DSFA	Development set forecast accuracy
DSO	Distribution system operator
DUoS	Distribution Use of System

E

E/P	Energy to power
EFR	Enhanced frequency response
ELM	Extreme learning machines
ERCOT	Electric Reliability Council of Texas
ERS	Emergency response service
ESFA	Evaluation set forecast accuracy
EU	European Union
EV	Electric vehicle

F

FCDM	Frequency control by demand management

FERC	Federal Electricity Regulatory Commission
FFR	Firm frequency response

G

GBT	Gradient boosted trees

H

HPO	Hyperparameter optimisation
HVAC	Heating, ventilation and air conditioning

L

LAD	Least absolute deviation
Li	Lithium ion
LM	Levenberg–Marquardt
LMP	Locational marginal price
LS	Least squares

M

ML	Machine learning
MAE	Mean absolute error
MAPE	Mean absolute percentage error

N

NG	National grid
NPV	Net present value

P

PJM	Pennsylvania – New Jersey – Maryland Interconnection

R

RBF	Radial basis function
RZM	Red zone management

S

SGD	Stochastic gradient descent
SoC	State-of-charge
SVM	Support Vector Machines

Chapter 1

Introduction

1.1 Background

"*It is a capital mistake to theorise before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts*", Sir Arthur Conan Doyle (1891). Over a century later, with the developments in technologies such as sensing, communication, computing and internet, large volumes of data are being generated and we have only recently started to resonate with this thought. The term data-driven is being used in contexts where data is at the core of decision making.

According to F. Rousseaux [1], data-driven thinking "*mobilises data to make connections or analogies, aiming to create certain configurations or to anticipate situations that might cause delays or predictive challenges*" and "*their performances are measured more by their ability to predict or to discover rather than to understand, explain or theorise*". Certainly, data-driven thinking is replacing the classical thinking that was based on theories. Within the domain of artificial intelligence, machine learning (ML) algorithms are being designed to build data-driven models that help gain insights from data. The ML algorithms may be supervised where the machine learns through example data or unsupervised where the machine learns on its own from random data. Today, data-driven models are being developed and deployed in domains such as engineering, finance and life sciences, among many others.

In the electricity sector, the legacy grids are being transformed to smart grids that enable the flow of information along with energy. As a result, there is increased availability of timestamped and high resolution data from the generation, transmission, distribution and consumption sides of the grid. These include: 1)

measurements, fault records and event records based on intelligent electronic devices; 2) power quality recordings at 50-60 times per second with geo-location based on phasor measurement units; and 3) electricity generation and consumption data from smart meters; among others [2].

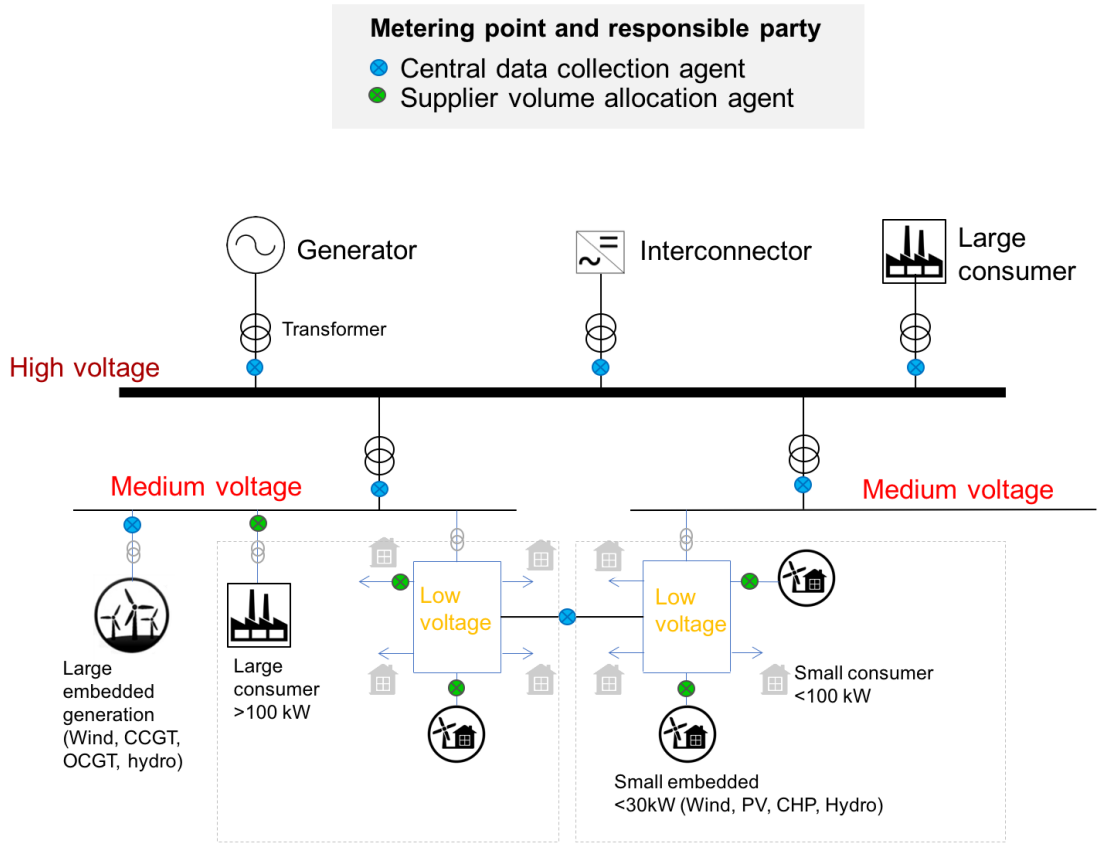

Figure 1: Metering data points in the UK electricity grid

Figure 1 shows the metering data points in the United Kingdom (UK) electricity grid that measure the flow of electricity. These include: 1) the transmission connected (high voltage) generators (coal, gas, nuclear, hydro, wind, solar), interconnectors (high voltage DC lines from Ireland, Netherlands, France) and large consumers (loads ≥ 100 kW); and 2) the distribution connected (medium/low voltage) embedded generators (wind, gas, hydro, solar), large consumers and small consumers (loads < 100 kW). Smart metering with communication capability is available for all the connected entities with the exception of small consumers and

small embedded generation (< 30 kW), which are aspiring to have them installed in the near future. This means that a large amount of generation and consumption metering data is already available in real-time, usually accessed as half-hourly datasets [3]. Although availability has increased, the use of such data for data-driven decision making is limited in the day-to-day functioning of the grid. This problem is elaborated further, in references to demand response (DR), which is an integral mechanism in the present-day smart grid operations.

1.2 Significance of data-driven modelling in demand response

This section briefly introduces DR to help describe the significance of data-driven modelling. Chapter 2 provides detailed discussions on the DR mechanism and its many advantages.

Grid balancing services work towards the common goal of maintaining electricity supply and demand in good balance. In the legacy grid, most of the balancing actions were provided by the centralised generators. Since communications to the demand-side were unidirectional, consumers were passive actors in the electricity system. The smart grids of today facilitate bidirectional communication with the consumers and they can actively participate in the operations of the grid through DR programs. Such programs enable consumer energy assets such as *flexible loads*, *standby generators* and *storage systems* to add value to the electricity system by offering grid balancing services from the demand-side. This is represented in the infographic in Figure 2.

The need for DR is warranted by the increasing grid imbalances due to intermittent renewables in the electricity generation mix. According to a recent study by the Rocky Mountain Institute, DR based balancing of renewables has the potential to avoid $1.9 billions of annual gas-fired generator costs and 20% of annual carbon emissions in the United States (US) [4]. Further, the impending roll-out of electric

vehicles and the ad hoc charging requirements will intensify grid imbalances. These could potentially be minimised through DR [5].

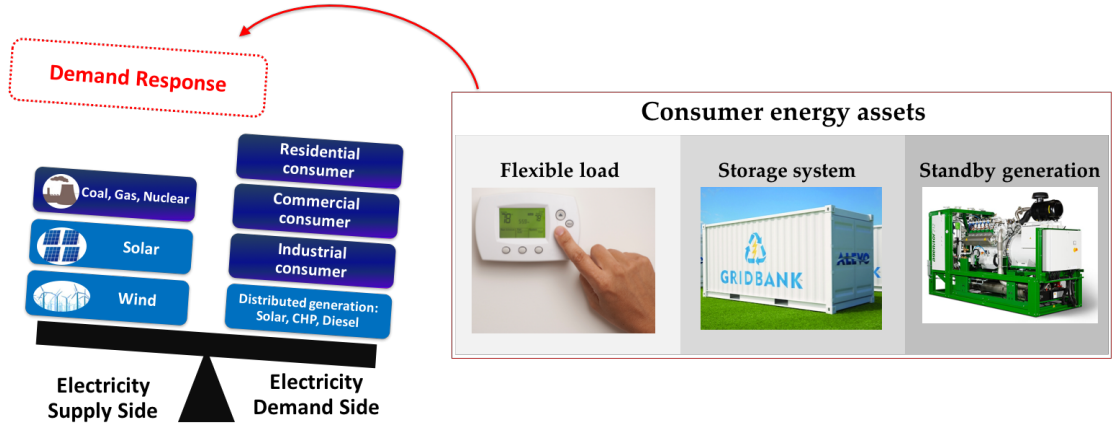

Figure 2: Consumer energy assets help mitigate the grid supply-demand imbalance through DR programs

Depending on the capacity and availability of their energy assets, residential, commercial and industrial (C&I) consumers participate in DR programs, either individually or through aggregators. The aggregators play an important role in DR with their capability to bundle small capacities of individual energy assets and deliver them to the grid as a single large energy asset. The different consumer energy assets such as flexible loads, standby generators and storage systems are summarised below.

Among the consumer loads, flexible loads are those non-critical loads that may be curtailed or shifted for a given duration of time without causing disruption to the consumer processes or comfort. Standby generators are installed to provide backup to the consumer loads during occasions of grid outages. Since outages are rare, these assets may be used as distributed generators for the grid. Compared to the centralised generators, operating distributed generators are cost-effective, especially during peak demand periods. Considering the transmission losses from centralised generation, distributed generation closer to the consumption points are more energy-efficient. Storage systems such as batteries or flywheels are usually installed

to maintain the power quality for critical consumer loads. In large-scales, they can store surplus energy from the grid and release when required, hence offering valuable services to a renewables-rich grid. The use of battery energy storage system (BESS) for DR will have implications on their life and costs (capital, operational and maintenance). Hence, proper energy management is required to ensure their continued operation and service delivery. Some of the DR related tasks based on these energy assets are summarised below.

Consumers responding to price-based DR programs (more discussion on different types of DR programs in Chapter 2) directly benefit out of the savings in their bills. For example, curtailment of flexible loads during peak price periods can help the consumer avoid those charges. In incentive-based DR programs, the value of the consumer energy assets to the grid is measured based on their availability and utilisation. DR related tasks such as *capacity scheduling*, *reliable operation* and *performance evaluation* are important in this context. Capacity scheduling enables the grid to estimate the availability of demand-side resources for system planning. For example, a known quantity of demand-side load curtailment capacity helps plan in advance for reduced generation. Grid contingencies due to unavailability of the scheduled capacity could be avoided through reliable operating mechanisms. For example, a consumer BESS with limited energy capacity could continue delivering a service if the appropriate energy management techniques are adopted. Performance evaluation helps assess the utilisation of an energy asset with reference to its committed availability. Under or over delivery of the scheduled capacity could be penalised, since it affects the grid stability. Performance evaluation also helps weed out the bad performers. Although there exist many other DR related tasks, the thesis specifically addresses capacity scheduling, reliable operation and performance evaluation of consumer energy assets.

A major share of the energy consumed worldwide is accounted for by large consumers such as large consumer buildings [6]. Few examples of large consumer buildings are shown in Figure 3. The research in this thesis focuses on DR from such

buildings and their energy assets such as flexible loads, standby generators and storage systems. These energy assets are part of the building processes and hence entwined with the whole building load. As a result, DR related tasks such as capacity scheduling, reliable operation and performance evaluation for these energy assets are dependent on the estimation of building load.

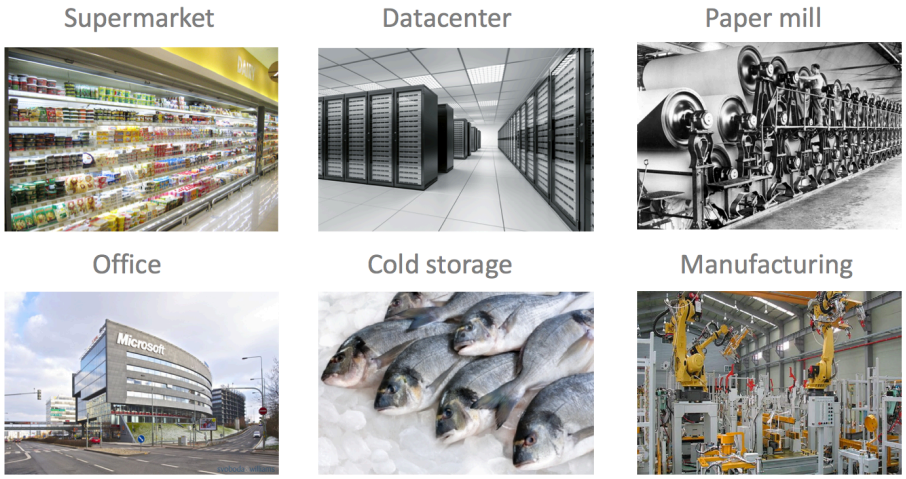

Figure 3: Examples of large consumer buildings

Physics-based or white-box models using inputs such as weather, geographic location, building design, building material properties, load characteristics, occupancy status and operating schedule, among others, provide the most accurate building load estimates [7]. Although DR aggregators facilitate the participation of building consumers in DR, not all information required to build physics-based models are made accessible to them due to privacy and security concerns. Data-driven models based on ML algorithms adopt a black-box approach for building load estimation where the historic load patterns are associated with accessible variables such as time-of-day and weather. This makes it easy to build reliable, replicable and cost-effective building load estimation models, particularly for DR applications.

Further, data-driven models are adaptable to changes in building conditions such as the addition of a new load even without the need for physical inspection or significant upgradation of the model. Adaptability and minimal intervention would mean that data-driven models are appealing for deployment in continuous DR operations. This would eventually encourage increased automation of DR activities.

As part of the smart grid transition, high resolution and real-time 'big data' related to building energy consumption, weather, system demand, generation, emissions and electricity prices are already available. Also, the domain of ML and data analytics is making immense progress. Hence, it is easy to collate the required data, develop ML based data-driven models and deploy them in DR related activities. In addition to this, data-driven models using minimal computational resources (developed in Chapter 4) are useful in internet-of-things (IoT) platforms that support large-scale deployment of connected devices on consumer sites with DR potential.

In the wake of increasing renewable generation and the impending electric vehicle deployment, there is a need for more flexibility in the grid through DR programs. Consumer behaviour modelling for DR is noted to be one of the major challenges for smart grids [8]. When larger number of consumer energy assets participate in DR programs, reliable, adaptable, cost-effective, computationally efficient, deployable and replicable models are required for many DR related tasks. In comparison to the physics-based approach, data-driven models are more promising in this regard.

On top of the direct benefits derived from data-driven modelling, creating pipelines of data pre-processing, model development and deployment, as visualised in Figure 4, would encourage the use of additional smart grid data and generate value out of them. For example, real-time emissions data from generators already available from the grid may be used to prompt the consumers for load curtailment using their flexible loads or storage systems during high emission periods. Such an emissions-based DR (demonstrated in Chapter 7), being experimented in some of the

electricity markets, would result in large emissions reduction on aggregated scales by shifting consumption away from carbon-intensive generation.

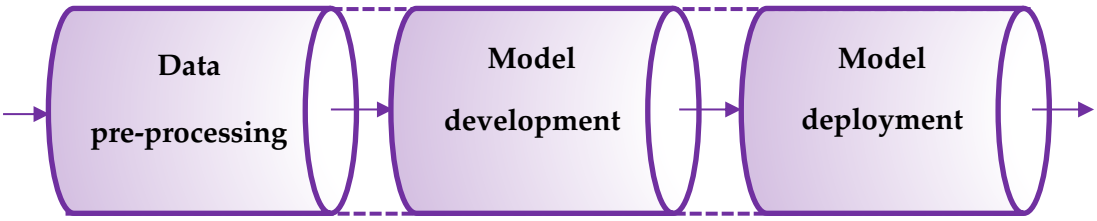

Figure 4: Data-driven pipeline considered in the research

In the long run, data-driven thinking would uplift DR as a more reliable and significant resource for the grid. Apart from DR, data-driven models can be used for better demand-side management. This includes identifying energy use patterns and suggesting energy efficiency measures for consumers. Data-driven models can also be used to forecast renewable generation, system demand peaks and imbalance prices. It is expected that, through data-informed thinking and better use of available data, the capabilities of artificial intelligence will seep into different layers within the electricity system operations. This is particularly important when the system is aspiring to be increasingly digitised.

1.3 Thesis statement

"Data-driven models utilise smart grid data and enhance the demand response from large consumer energy assets such as flexible loads, standby generators and storage systems"

1.4 Aims and objectives

Data from smart grids are increasingly available but seldom utilised to improve the electricity system operations. The thesis aims to critically assess the possible use of smart grid data in facilitating demand response (DR) from large consumers.

The main objective of this research is to develop a data-driven model based on machine learning (ML) for large consumer building load estimation and to provide conclusive evidences of its capability to improve DR operations. A secondary objective of this research is to explore how smart grid data such as real-time grid emissions can be used to operate an emissions-based DR program to shift electricity consumption away from carbon-intensive sources.

The specific research objectives are:

- to develop a reliable, computationally efficient and deployable data-driven model for large consumer building load estimation based on the approaches adopted in modern data science.
- to apply the data-driven model in DR operations such as capacity scheduling and performance evaluation of large consumer buildings (with load curtailment capability using flexible loads and standby generators) and demonstrate the improvement over conventional models employed in ongoing DR programs.
- to use the data-driven model for development of a novel state-of-charge management strategy that enables the reliable operation of a large consumer battery energy storage system in multiple DR programs.
- to assess the emissions reduction potential of large consumers participating in a conceptual emissions-based DR program designed based on the real-time grid emissions data and available load flexibility.

1.5 Scope

The scope of the thesis is limited to the following:

- Smart meter data is available for large consumers because of the favourable regulations in most electricity markets. The data-driven building load estimation model developed in this thesis is demonstrated for large consumers such as commercial buildings. Data-driven models could be developed for industrial

buildings with predictable load patterns as well, although not demonstrated. For industrial buildings such as manufacturing plants with random or unpredictable load variations, the model may not be able to offer good performance without including relevant causal data.

- Smart meter roll out for residential consumers is either in the initial stages or incomplete in most electricity markets. Even in advanced electricity markets such as the UK, most of the residential electricity consumption is billed based on estimated load rather than metered load. Further, participation of residential consumers in the DR programs face restrictions in these markets. Hence, the data-driven modelling work presented here does not address the residential consumers. However, the adopted methods may be tested for these small consumers with appropriate alterations. Considering the stochasticity in their consumption behaviour, it would be interesting to identify the influencing factors and use them for data-driven modelling. Given that sufficient data are available, such a bottom-up modelling approach from individual residential loads to aggregated levels would enable better DR.
- During the development of the data-driven model, a heuristic optimisation method is prioritised over other methods, due to its computational efficiency and ease of replication. In the future, if more reliable and computationally efficient optimisation methods are available, they can replace the heuristic method adopted in the model.
- The DR related tasks such as capacity scheduling, reliable operation and performance evaluation considered in the thesis are in fact a subset of the many system operational activities. Data-driven modelling could be applied for other tasks that involve consumer building loads such as recognising energy use patterns and suggesting energy efficiency measure. However, such applications are not modelled in this thesis.
- Ongoing DR programs and their frameworks have been used to demonstrate the real-world operational issues and also to make comparisons of the conventional models in use with the data-driven models developed in the thesis. It is to be

noted that, due to the volatile nature of this industry, the DR programs and their requirements are susceptible to changes from time to time.

- The emissions reduction potential is estimated for large consumer buildings in selected US system operator regions using a publicly available grid marginal carbon emissions intensity data derived using the LEEM 2.0 method (refer chapter 7). There is no consensus yet, on the best emissions estimation method used to derive such data. However, the reliability of this method is assured based on multiple consultations between researchers, industrial partners and energy regulators in the US. The emissions reduction potential of residential buildings is not estimated in the thesis, although they are expected to be a major consumer segment if an emissions-based DR program is launched in the future.

1.6 Contributions

The presented research was largely conducted in an industrial setting at a DR aggregator company, Flexitricity Limited (Edinburgh), one of the industrial partners in the *Advanced communications and information processing in smart grid systems (ADVANTAGE)* project. Flexitricity has a portfolio of large consumer buildings participating in different DR programs in the UK. A brief part (four months) of the research was also conducted at the Rocky Mountain Institute (Colorado), one of the pioneers in DR research.

In the presented thesis, Chapters 4—7 are considered as the core chapters. The link between these chapters and their placement in the thesis are represented in Figure 5. Chapter 4 develops a data-driven building load estimation modelling approach and this is applied in different DR operational contexts in Chapter 5 and Chapter 6. Chapter 7 takes a detour and highlights the implication of setting up data-driven pipelines in facilitating the use of available smart grid data. The detailed layout of the thesis is discussed in the next section.

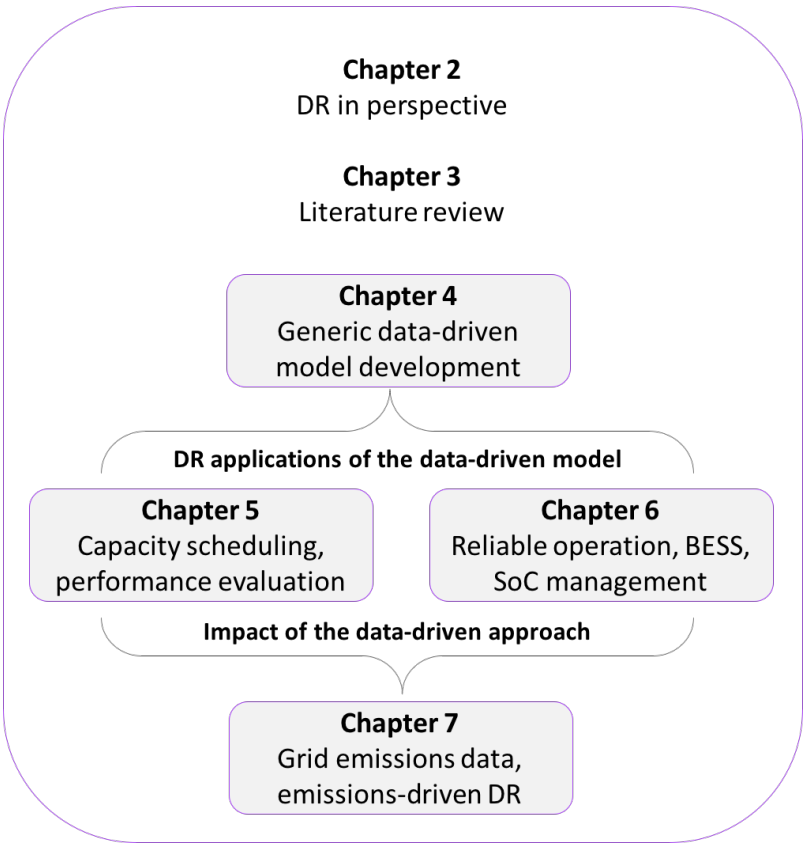

Figure 5: Link between core chapters and their placement in the thesis

The main contribution of the thesis is a critical evaluation of the use of smart grid data for enhancing DR from large consumers. The contributions of the core chapters (4 –7) are discussed below.

A data-driven modelling approach for large consumer building load estimation is developed using supervised ML in Chapter 4. This is performed based on the structured modelling approach adopted in modern data science giving great attention to different stages of the data-driven pipeline such as data pre-processing and model development. Such approaches are missing in the previous building load estimation studies. In the course of this modelling work, certain gaps between empirical and applied machine learning have also been identified. For instance, good practice in supervised ML recommends the comparison of predictive performances of different ML algorithms after calibration using hyperparameter

optimisation (HPO), which has been found to be missing in certain building load estimation studies.

Computational efficiency without compromising the model's reliability is considered as an important design criterion in the modelling work. For example, beyond their general applicability on building load data, selection of ML algorithms is motivated by their computational resource requirement. Similarly, a random-search HPO has been adopted in model calibration for its judicious use of available processing power and the potential to perform multiple rounds of calibration instead of a computationally demanding grid-search HPO method. Despite its advantages, random-search has not been widely adopted in ML based building load modelling.

The modelling process has prioritised aspects related to deployment of the data-driven model in an industrial operational environment. This has also influenced the selection of methods in the data pre-processing and model development stages. For instance, data collection focuses only on the predictor variables that are realistically accessible for a DR aggregator. Feature selection tries to minimise the use of predictors with inherent errors, such as weather forecasts and their derivatives. Cross-validation is performed based on a forward sliding window of training-testing sets, simulating the actual training and forecasting process after deployment. The study also experiments the effect of possible data gaps between the training-testing sets on the model performance. Development of such deployment centric data-driven building load estimation models are limited in the literature.

Setting up a data-driven pipeline involving data pre-processing, model development and model deployment stages is important to ensure that the modelling is replicable on a large number of buildings. This is particularly important in these times when the need for DR capacity is higher in the grid and the participation of large building consumers is ever increasing. However, building load modelling studies focussed on replicability of the methodology for large-scale applications are limited. The work in Chapter 4 fills this gap by developing a

modelling methodology that is applied in different DR related activities in Chapter 5 and Chapter 6.

Chapter 5 discusses application of the data-driven building load estimation model developed in Chapter 4 for capacity scheduling and performance evaluation. These DR operational tasks are specific to the incentive-based DR programs. Only a few previous studies have explored the benefits of data-driven models in such contexts. Further, these studies have seldom focussed on the large-scale deployment of data-driven models by making them computationally efficient and replicable, as done by the presented study. Different types of large consumer buildings such as a supermarket, a laboratory, a hotel, an office, a retail store and a hospital have been used for the load estimation modelling. The accuracies of supervised ML models for load estimation in these buildings are shown to be better than that of the conventional models used in the ongoing DR programs. Large-scale deployment of such data-driven models in DR operations can increase the reliability of capacity scheduling and performance evaluation, benefitting the grid as well as the building consumers.

Chapter 6 discusses application of the data-driven building load estimation model developed in Chapter 4, for reliable operation of a battery energy storage system (BESS) in a large consumer building site. State of charge (SoC) management is important for the continued operation of a consumer BESS participating in DR programs. Most of the existing SoC management strategies found in the literature are rule-based and these may not always be beneficial for the BESS owner or for the grid in practical DR operational scenarios. For instance, when a large number of consumer BESS units participating in an evening peak shaving DR program perform scheduled recharging for SoC management, they would burden the grid and cause further imbalance. Also, the BESS owners have to pay for this additional energy consumed. In this study, the day-ahead building load forecasts based on the modelling discussed in Chapter 4, enable the implementation of a data-driven SoC management strategy without affecting the grid stability and at the same time not

adding to the BESS operational costs. The data-driven SoC management strategy is developed based on the framework of ongoing multiple DR programs. The economic viability of a university building BESS participating in the DR programs operating based on the developed strategy is assessed through a case study. To the best of our knowledge, a data-driven SoC management strategy has not yet been studied in the literature and this could motivate further research in data-driven modelling for consumer BESS based DR.

In most of the electricity markets, conventional price/incentive-based DR programs focus only on grid balancing. Chapter 7 contributes to the concept of an emissions-based DR program targeted at grid emissions reduction, through load curtailment of large consumers, making use of the real-time grid emissions data from the smart grid. This is particularly interesting for consumers with emission reduction targets and social responsibility obligations. The study provides a detailed assessment of available flexibility in large consumer buildings in the US and estimates their emissions reduction potential for different realistic scenarios. The chapter highlights the possible use of available but under-utilised smart grid data based on data-driven pipelines that can be easily replicated. It also underscores that, an emissions-based DR program can open up the DR market for large consumers who may not otherwise be interested in the conventional DR programs.

Apart from the chapter specific contributions discussed above, the thesis also makes the following general contributions.

Based on the extensively surveyed literature in [9,10], DR is usually formulated as an optimisation problem dealing with objectives such as social welfare (utility) maximisation, cost minimisation, energy consumption minimisation or combinations of these. The studies have also addressed issues such as the impact of prices, renewables, storage and electric vehicles in the DR paradigm. However, research focused on problems associated with ongoing DR programs are limited. The research presented here attempts to fill that gap.

The industrial setting for this research has allowed access to building load data from diverse large consumers that are usually inaccessible for academic research. Variants of the data-driven model developed in this research have proven useful for deployment in the ongoing DR operations at Flexitricity. These models promise to advance the automation strategies in DR operations that are presently semi-automated or manual.

Flexitricity has also proposed the data-driven modelling developed in this thesis to the *Department for Business, Energy & Industrial Strategy (gov.uk)* in their *Innovative Non-Domestic Demand-Side Response Competition*. The project aims to identify, test and disseminate learning from innovative approaches to DR in operational, non-domestic applications in the UK. As part of this project, the building load estimation models will be deployed on a wider scale to tap into the demand-side flexibility from large consumers.

The results of the emissions-based DR study were discussed in a multi-stakeholder workshop in Chicago, organised by the Rocky Mountain Institute. As of today, the emissions-based DR is being conceptualised and experimented in electricity markets such as the US and the European Union (EU) towards encouraging more renewable energy consumption. This may help increase the DR participation from building consumers, adding more energy assets to the flexibility pool.

1.7 Layout

The thesis is structured into a total as eight chapters. These are summarised below.

Chapter 2 provides a holistic perspective on DR and builds the background on which rest of the chapters could be better comprehended. The chapter introduces a definition of DR and examines it within the scope of smart grids. The different DR concepts of peak shaving, valley filling, etc. are discussed along with parameters commonly used in DR programs. The physical implementation of DR from

consumer energy assets is covered in this chapter. DR programs in different electricity markets and their present status are also discussed.

Chapter 3 reviews the literature that aligns with the concepts and motivation of the research presented here. Building load estimation models are reviewed and evaluated with respect to the DR application requirements. Supervised ML algorithms and model calibration methods applied in building load estimation studies are discussed. General data-driven models developed for DR applications such as capacity estimation and performance evaluation are reviewed and gaps are identified. State-of-the-art state-of-charge (SoC) management strategies for reliable operation of a consumer BESS in DR programs are reviewed and the need for a data-driven strategy is highlighted.

Chapter 4 develops a data-driven modelling approach focussed on reliability, computational efficiency and the potential for large-scale deployment. Different stages such as data pre-processing, model development and model deployment are addressed in detail. Diagnostics and demonstrations are performed using an office building dataset. Day-ahead and week-ahead building load estimations are developed using three different supervised ML algorithms. The data-driven modelling approach developed in this chapter is used for specific DR operational tasks such as capacity scheduling, performance evaluation and reliable operation in the succeeding chapters.

Chapter 5 applies the data-driven building load estimation model in the 1) capacity scheduling of a supermarket, a laboratory and a hotel building, and 2) performance evaluation of a retail store, an office and a hospital building. These large consumer buildings are assumed to have load curtailment capability based on standby generators or flexible loads. Applications of the data-driven models are demonstrated based on an ongoing frequency responsive DR program in the UK and an emergency reserve DR program in the US electricity markets. These models are compared with the conventional models used in the different electricity markets and are shown to perform better.

Chapter 6 develops a novel data-driven SoC management strategy based on the building load estimation model for reliable operation of a large consumer building BESS in multiple DR programs. A frequency responsive DR program is considered along with a peak shaving DR program in the UK electricity market. The SoC management allowance provided by the frequency responsive DR program is used to recharge the BESS every day for providing load backup to a building during evening system peak period (red-zone). This recharge set-point is informed by the data-driven building load estimation model one day in advance. A case study is performed to simulate the data-driven SoC management strategy on a university building BESS. Further, a sensitivity analysis is used to prove the robustness of this method, followed by an economic evaluation of the project.

Chapter 7 explores the emissions reduction potential of large consumer buildings based on an emissions-based DR program. The smart meter data is collected from buildings that are already participating in conventional DR programs in the US. The understanding of available load flexibility is used to build scenarios for estimating the emissions reduction potential. For consumers that do not participate in conventional DR programs, emissions-based DR program is proposed as an alternative route to the DR market.

Chapter 8 concludes the work presented in the thesis. A summary of the main findings of the research is given in this chapter. The activities pursued for knowledge dissemination as part of the research are reminisced. The specific and broader impact of the presented work are discussed. Following this, the limitations of the research and the future scope of work are also stated.

Chapter 2

Demand response in perspective

2.1 Introduction

Consumers play an important role in the functioning of electricity systems from the demand-side. Demand-side management (DSM) is understood as a set of actions to plan, monitor and control the energy consumption on the demand-side. Energy efficiency and demand response (DR) are the two vehicles of DSM implementation [11]. This chapter provides a holistic perspective on DR and lays the foundation for rest of the thesis.

The Federal Electricity Regulatory Commission (FERC) in the United States (US) defines DR as: *"changes in electric use by demand-side resources from their normal consumption patterns in response to changes in the price of electricity, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized"* [12]. This definition of DR is adopted in the thesis.

According to the above definition, DR mechanism elicits response from the demand-side based on electricity prices or incentives towards mitigating grid contingency events. DSM actions such as DR could only be realised if the electricity grid supports the flow of information along with energy. This identity of an electricity grid as a smart grid is illustrated in Figure 6. It is represented that, apart from the energy flow between generation, transmission and distribution segments of the physical grid, information flow between different entities is equally important. Smart grids are differentiated from the legacy grids based on the amount

of sensing, control and communication technologies used in system operations. DR is an integral aspect of the smart grids enabled by such technologies.

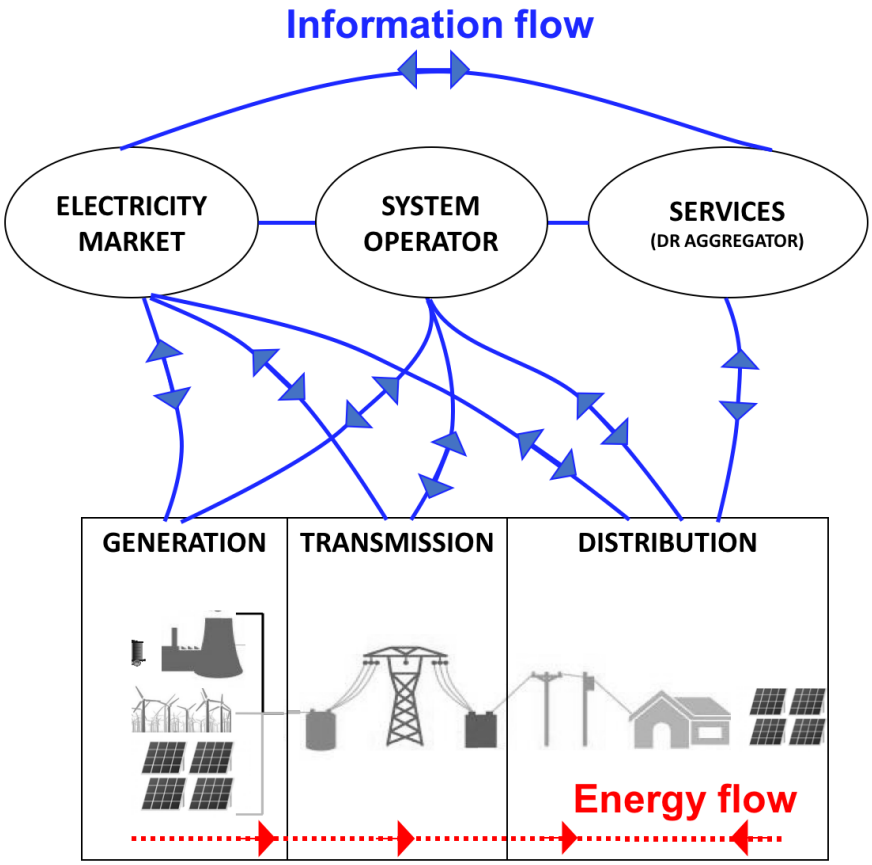

Figure 6: Information and energy flow in the smart grid (adapted from [13])

The rest of the chapter is structured as follows. Section 2.2 discusses the different types of DR actions. The benefits of DR are given in Section 2.3 . The terminologies used in the DR industry are elaborated in Section 2.4. Section 2.5 considers the physical implementation of DR. Finally, Section 2.6 reviews the ongoing DR programs in different electricity markets.

2.2 Types of DR actions

In DR, the changes in electricity consumption on the demand-side could be categorised into the following four categories: peak clipping, valley filling, load

shifting and dynamic energy management [14]. These are demonstrated based on consumer load shapes in Figure 7. Peak clipping or shaving refers to load curtailment during periods of system demand peak. This response from consumers may be induced by higher electricity prices. Valley filling considers the increase in consumer load during system off-peak demand periods, based on lower electricity prices. Electric vehicle (EV) charging is a potential candidate for valley filling. Load shifting is a combination of peak clipping and valley filling. Dynamic energy management refers to the use of consumer flexibility in continuous response to electricity price or other signals from the grid.

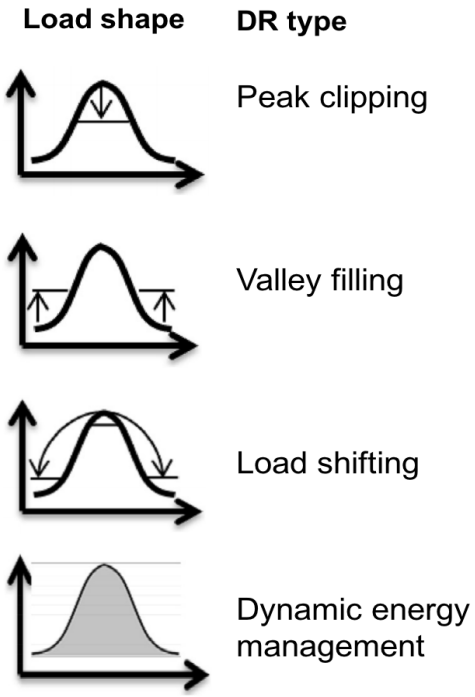

Figure 7: DR types represented on load shapes [14]

2.3 Benefits of DR

DR from consumers has multifarious benefits. These are discussed below with examples.

Figure 8 presents a simulated solar and wind only generation scenario compared with the actual national demand in the US electricity system for the year 2015. Such high imbalances call for more balancing resources in the electricity systems with high renewable mix. Using marginal generators may not be sufficient to meet these requirements. Consumer energy assets such as flexible loads, standby generators and storage systems promise to provide the required flexibility to solve this impending issue through DR programs. Particularly, consumer BESS can provide energy arbitrage, where stored excess energy is sold during high demand periods [15] as well as capacity firming, that can smoothen the wind plant output and avoid power swings.

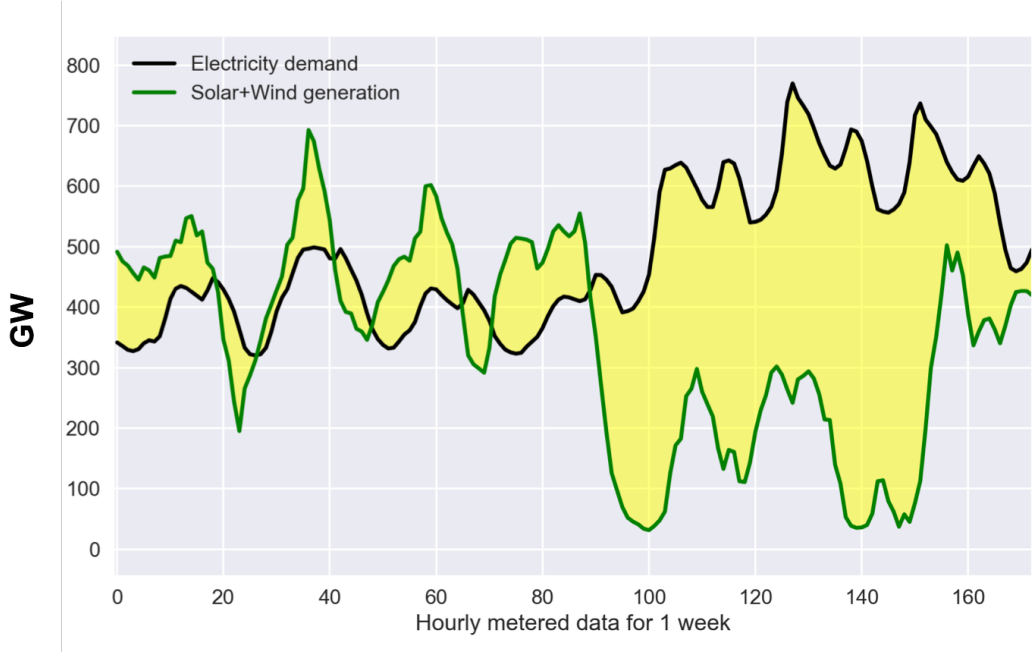

Figure 8: Demand-supply imbalance simulated in a solar and wind only generation scenario in the US electricity system (reference year is 2015)

Figure 9 shows the uncertainty in wind generation forecast across three consecutive days in the UK electricity system [16]. This introduces challenges in the system planning even for short-term, let alone long-term. Increased flexibility from consumer energy assets with sufficient speed of response can deal with short-term uncertainties in the grid, reliably.

Another issue with the increasing renewable energy penetration is the usage of power electronics converters and the subsequent lack of rotating mass (system inertia) in the grid [17]. This makes the grid more vulnerable to events such as generation failure. Fast responding (sub-second) frequency responsive consumer energy assets enable the grid to face such events without causing blackouts. Hence DR plays a key role in maintaining system reliability.

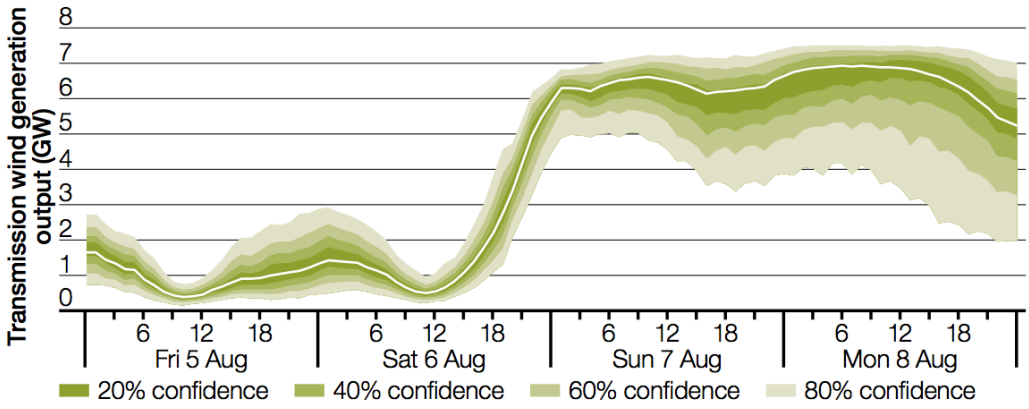

Figure 9: Uncertainty in the UK transmission system connected wind generation forecast [16]

The possibility of system demand peak shaving through load curtailment, defers the need for investment in grid infrastructure as well as marginal generators. DR programs introduce more competition in the wholesale electricity market by providing a cost-effective and low emission alternative to conventional generation [18]. The price – energy plot in Figure 10 shows the electricity supply and demand curves. Inelastic DR (D1), with no flexibility on the demand-side, results in an inelastic wholesale electricity clearing price (P1). Whereas, elastic DR (D2) utilising the demand-side flexibility enables reduction in the clearing price (P2) [19].

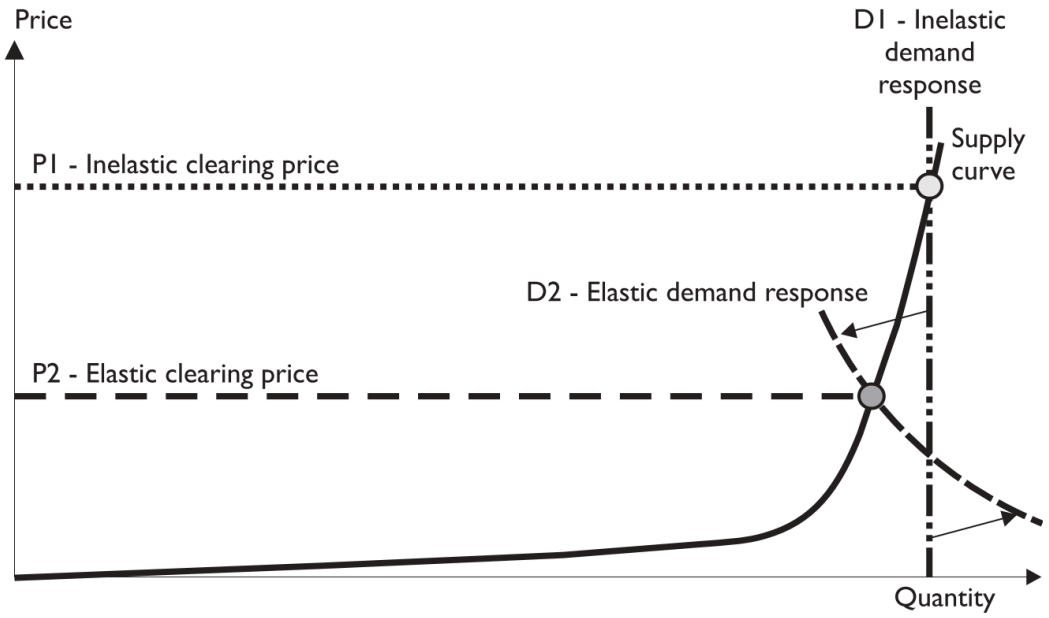

Figure 10: Impact of demand-side flexibility on wholesale electricity price [19]

Flexible loads such as HVAC enable load curtailment without affecting occupant comfort in large consumer buildings [20]. Hence, large consumers could use this flexibility to achieve savings in electricity bills during peak price periods. Many consumer energy assets such as standby generators are underutilised and their availability adds value to the grid while also enabling the owner to earn through DR programs. Storage systems within consumer premises could also provide load curtailment, hence earning additional revenue. All these factors can be tapped for an effective DR implementation.

2.4 DR industry terminologies

Some terminologies used in the DR industry are elaborated further.

2.4.1 DR operators and participants

The utilities or transmission system operators (TSOs) that operate DR programs in different electricity markets are referred to as DR operators. The utilities operate retail DR programs, whereas the TSOs operate the wholesale DR programs. The C&I

consumers may participate in the retail as well as the wholesale DR programs. However, residential consumers are still limited to participation in the retail programs. If consumers do not have the technical expertise to execute DR or their small energy assets do not qualify the minimum capacity requirement for participation, aggregators with a large portfolio of consumer energy assets play the intermediate role. The consumers and aggregators who participate in DR programs may be referred to as DR participants.

2.4.2 DR events and parameters

The events such as unforeseen generation failure, peak demand periods, surplus wind generation, etc., based on which DR programs are designed are referred to as DR events. Some of the common parameters used in the DR industry are listed below. These are also contextualised based on load curtailment response from a commercial building shown in Figure 11.

- ***Notification***: A DR event may occur anytime depending on the grid conditions. Notification of the DR event maybe offered as phone calls, emails or messages to the DR participant. Notification period is the time available for an energy asset to start responding to the notification.
- ***Trigger***: In the absence of notification, the energy assets are triggered to respond based on grid signals such as electricity prices or frequency.
- ***Response time***: The time taken by the energy asset to start delivering and reach the full committed capacity. Some of the energy assets like BESS can respond in sub-seconds while it might take a couple of minutes for an air-conditioning unit to be fully curtailed.
- ***Duration***: The period of time for which an energy asset provides response during a DR event. The duration may either be fixed or vary depending on the type of the DR event.
- ***Flexibility***: The full committed capacity that is curtailed or increased by the energy asset based on the requirement of the DR program is referred to as flexibility.

- ***Number of DR events***: The number of DR events may be given in the DR program contract on an 'events per month' or 'events per year' basis. Alternatively, the total expected duration of all DR events on an 'hours per month' or 'hours per year' basis may be given.
- ***Rebound***: For loads such as HVAC systems, the lost heating or cooling energy during a DR event is, in specific cases, recovered through rebound that results in an energy spike after the event.

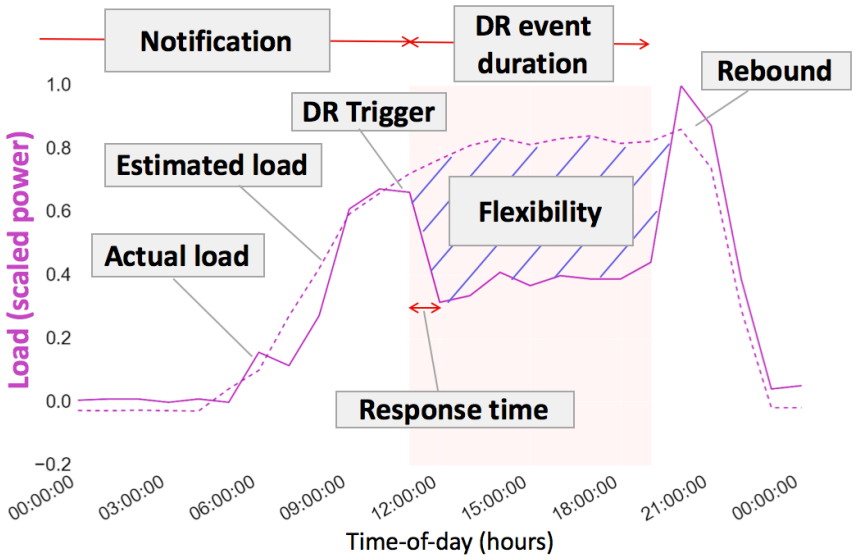

Figure 11: A representative DR event and parameters

2.4.3 Grid signals

The grid signals provided to DR participants so as to trigger response may include economic signal, reliability signal or even emissions signal.

- ***Economic signal***: Economic signals such as time-of-use (ToU) and dynamic electricity prices (shown in Figure 12) encourage consumers to perform price arbitrage by consuming less when electricity prices are high [21].

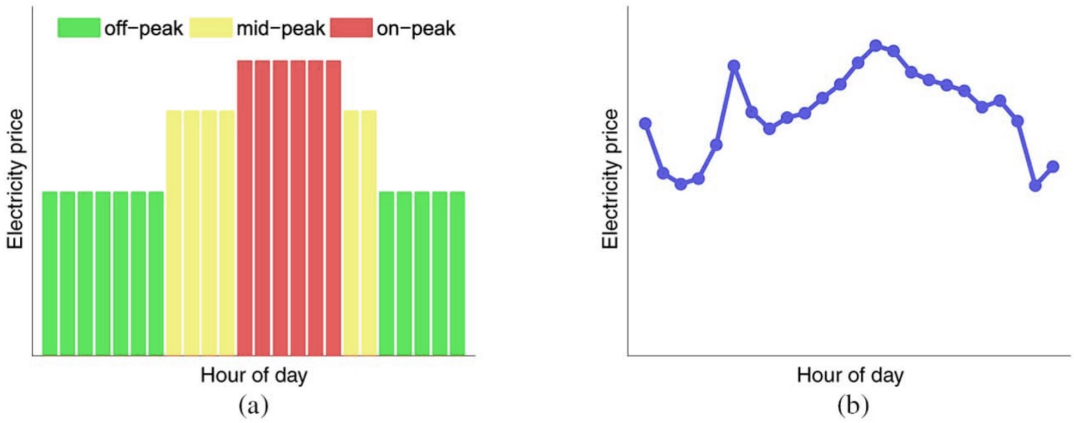

Figure 12: Economic signals such as: ToU price (a) and dynamic electricity price (b) [21]

- ***Reliability signal***: This primarily refers to the frequency signal that helps monitor the energy demand-supply balance in the electricity grid. The nominal frequency depending on the electricity system is either 50 Hz (UK/EU) or 60 Hz (US). The deficit of energy in the grid, pulls the frequency below nominal and excess of energy pulls it up. Small deviations from nominal frequency are balanced by the system inertia (providing by rotating mass of the generators). Large deviations occur when there is a generation failure or sudden change in demand [22]. In frequency responsive DR programs, the operational capacity from consumer energy assets are expected to: 1) respond at a pre-defined frequency threshold (static response shown in Figure 13), 2) vary proportionally with the frequency signals (dynamic response shown in Figure 14) or 3) vary with rate-of-change-of-frequency (RoCoF) signals also referred to as inertia response.

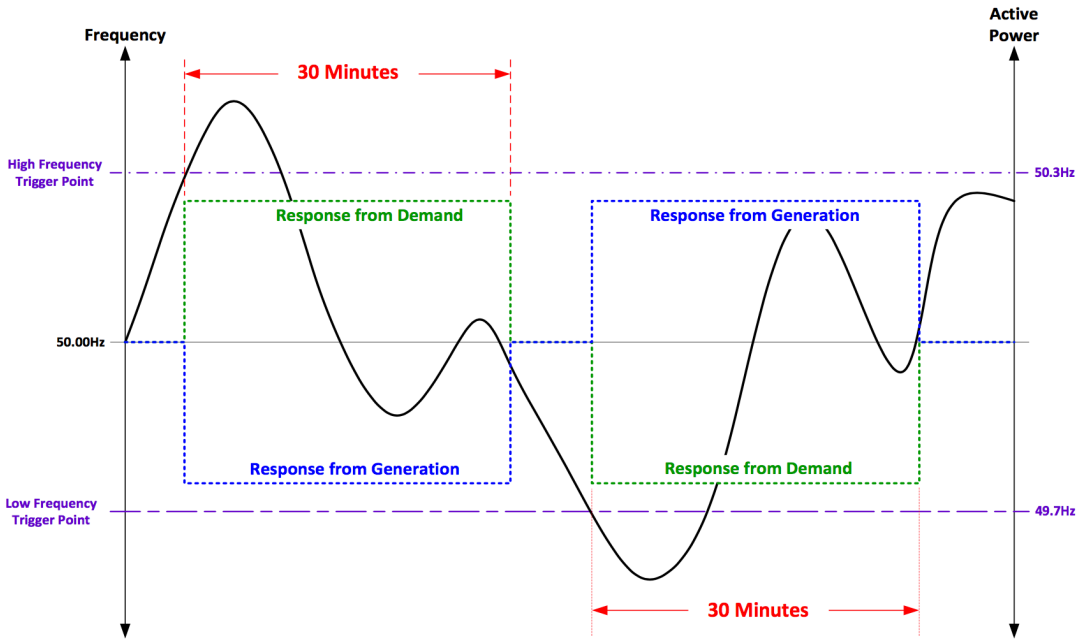

Figure 13: Static frequency response (black curve represents the grid frequency) [22]

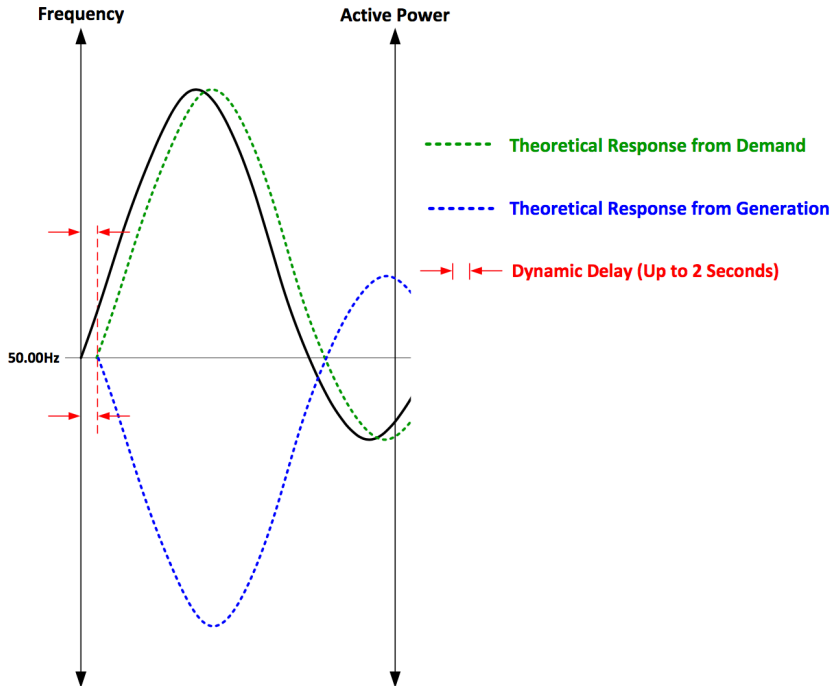

Figure 14: Dynamic frequency response (black curve represents the grid frequency) [22]

- ***Emissions signal***: Real-time grid carbon emissions data (from generators) such as that mapped in Figure 15 for Europe are publicly available. Although not

being practised as a DR program in any electricity market as of today, these grid signals could be used to trigger DR from consumer energy assets. In times of higher grid emissions due to fossil-fuel based generation, consumers may be encouraged to reduce demand and vice-versa.

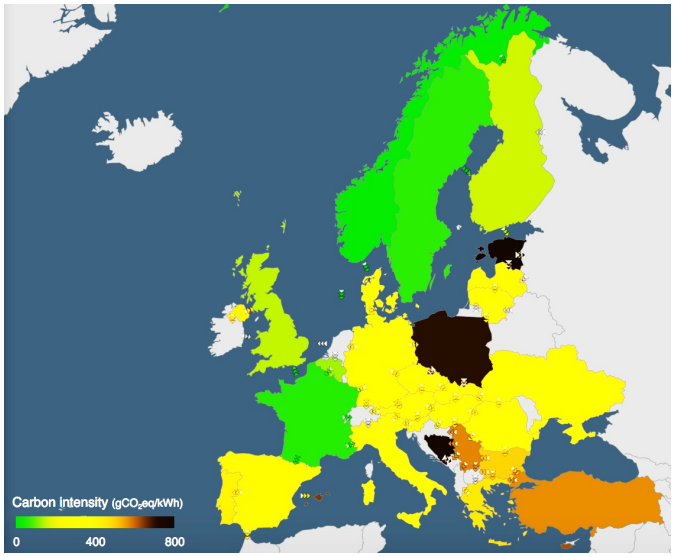

Figure 15: Real-time grid carbon emissions map of EU [23]

2.4.4 Motivation and rewards for participation

Based on the motivation, DR programs are conventionally classified as incentive-based and price-based [10]. Those consumers who are not interested in the rewards of these conventional DR programs may be motivated by the possibility of an emissions-based DR. These are summarised in Figure 16 and discussed below.

- ***Incentive-based DR:*** These programs motivate consumer participation by offering payments based on availability and/or utilisation of their energy assets. The availability payments are offered for the DR capacity scheduled for pre-defined time-periods referred to as windows. The utilisation payments may be offered for the energy delivered or consumed, as seen from the grid, during the DR event.

- ***Price-based DR***: These programs induce response from consumers based on economic signals. The inherent motivations for consumers to minimise their electricity costs is exploited in the price-based DR programs. The reduction in electricity costs is the direct reward for participation.
- ***Emissions-based DR****:* Similar to price-based DR, emissions signals also have the potential to induce response from consumers concerned about reducing their global warming footprint. The rewards maybe quantified in terms of carbon savings.

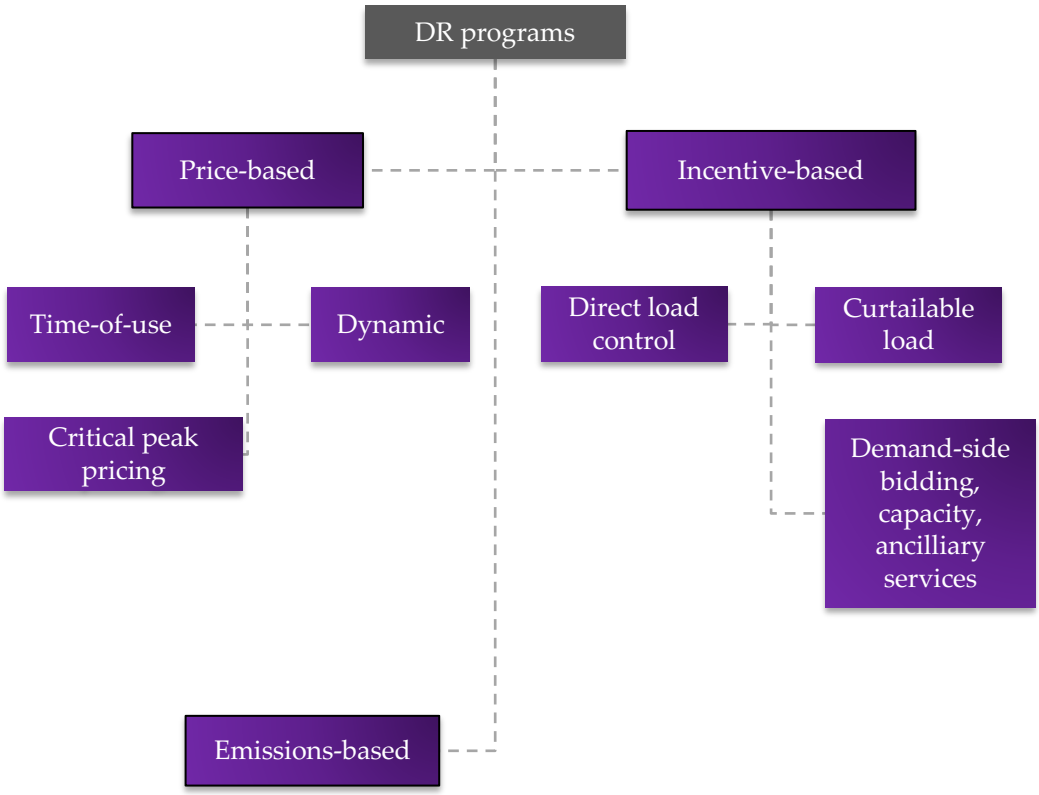

Figure 16: DR programs categorised based on motivation and rewards. Emissions-based DR has been considered as a new category in this thesis.

2.5 Physical implementation of DR

DR is physically implemented through the set of sensing, monitoring, control and communication technologies that are also part of the smart grid. A general layout

for physical implementation of DR is shown in Figure 17. Depending on the availability of the consumer energy assets such as flexible loads, standby generators or storage systems, as well as the consumer's technical capability, they may enrol for the DR programs directly with the grid (system operator or utility) or through an aggregator company. The sensing, monitoring, control and communication interface is an essential component for enabling DR. There is bidirectional communication between the grid, aggregator as well as the interface on the consumer site. The type of communication channel may vary across the use cases. For frequency responsive DR programs, grid signals such as frequency are usually detected on site through the interface at the consumer site. DR is provided directly to the physical grid based on the communication signals (notification) or grid signals. The success or failure of this response is reported back to the aggregator or grid through the communication channels.

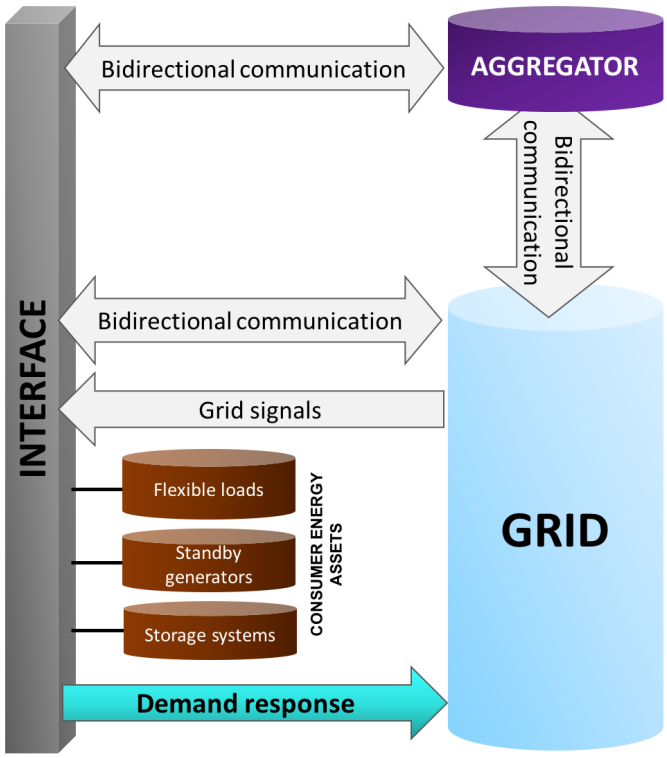

Figure 17: Physical implementation of DR

2.6 DR programs in different electricity markets

The discussion below is limited to electricity markets in the US, the UK and mainland EU. To learn more about DR programs in other markets, the following review is recommended: [5].

2.6.1 DR programs in the US

DR programs were introduced to the world for the first time by the US electricity market in the 1970's. These were operated by the vertically integrated utility companies for peak shaving when increased air-conditioning usage caused demand peaks. Since the deregulation in 1990's, the system operators (marked in Figure 18) operate the wholesale DR programs whereas the utilities within the system operator regions as well as the vertically integrated ones operate the retail DR programs. These programs are broadly categorised within the regulation, reserves, energy and capacity market services [24].

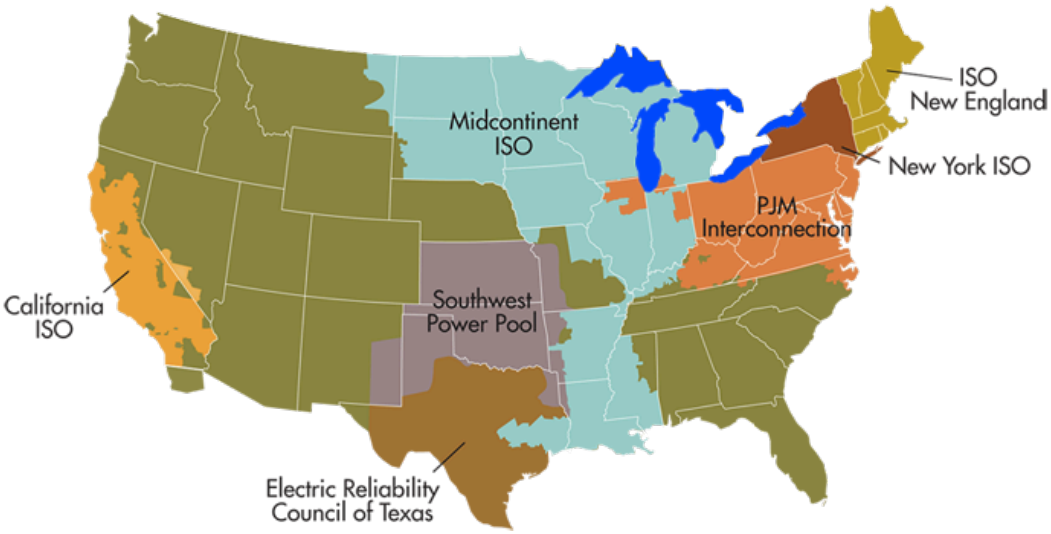

Figure 18: System operators in the US [25]

The regulation programs include frequency response services that provide the first level of support to grid during contingencies. Fast responding regulation service in ERCOT (Texas) is an example. Reserve programs provide the second level of

support after regulation services. Synchronised reserve service in PJM (Pennsylvania - New Jersey - Maryland) interconnection is an example. Energy programs enable consumer energy assets to offer load curtailment in the wholesale electricity market or as peak shaving services in the retail market. Capacity market programs ensure that sufficient capacity is bid from consumers before a shortfall season. Emergency DR program in PJM interconnection is an example [24]. Figure 19 shows these categories based on the duration and notification parameters with a few example DR programs. Apart from ERCOT and PJM, the listed system operators include NYISO (New York), CAISO (California) and ISONE (New England).

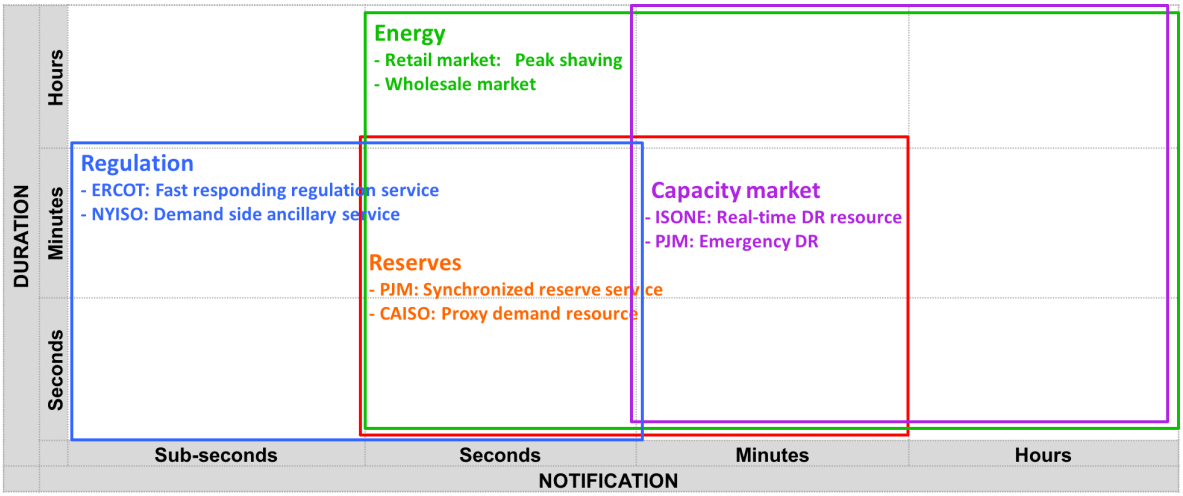

Figure 19: DR programs in the US categorised based on duration and notification parameters (example programs also included)

By 2015, large consumers such as commercial buildings contributed 56 GW of DR capacity. This was ~7% of the system peak demand. The tapped DR capacity was better than the 'business as usual' projection from 2009 as shown in Figure 20, demonstrating the success of DR programs in the US [26][27][28].

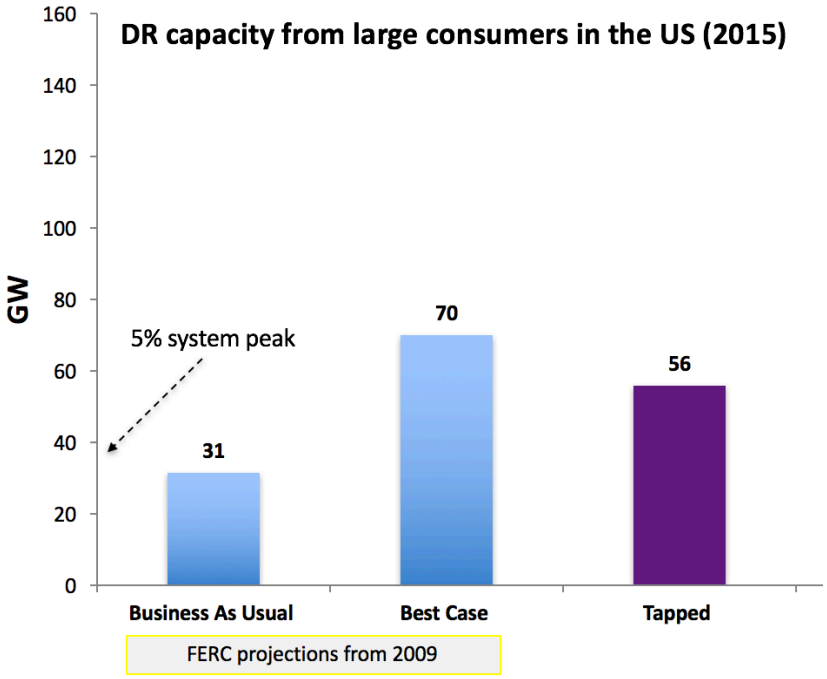

Figure 20: Business as usual and best case projections from 2009 are compared with the tapped capacity in 2015. Plotted using data collected from [27][26][28].

2.6.2 DR programs in the UK

The transmission system operator (TSO) National Grid (NG) operates the major DR programs in the UK; a few utility operated DR programs/pilots also exist. DR participation is allowed only for large consumers and their aggregators. The DR programs are broadly categorised into frequency response (same as regulation in the US), reserves, capacity market and peak management (energy). It is estimated that, over 2.7 GW of DR capacity (excluding BESS) was available in the first quarter of 2017 [29]. In addition, there is an improved regulatory interest in removing the barriers to DR program implementation. According to the future energy scenarios developed by NG [30] (shown in Figure 21) DR capacity in the UK is expected to increase, particularly in the 'two degrees' scenario where sustainability is the top priority. The influence of BESS adoption is also notable in these scenarios. This is supported by the fact that 4 GW of BESS has already been connected to the UK grid

as of today [31]. Based on [32], the ongoing DR programs in the UK electricity market are summarised further.

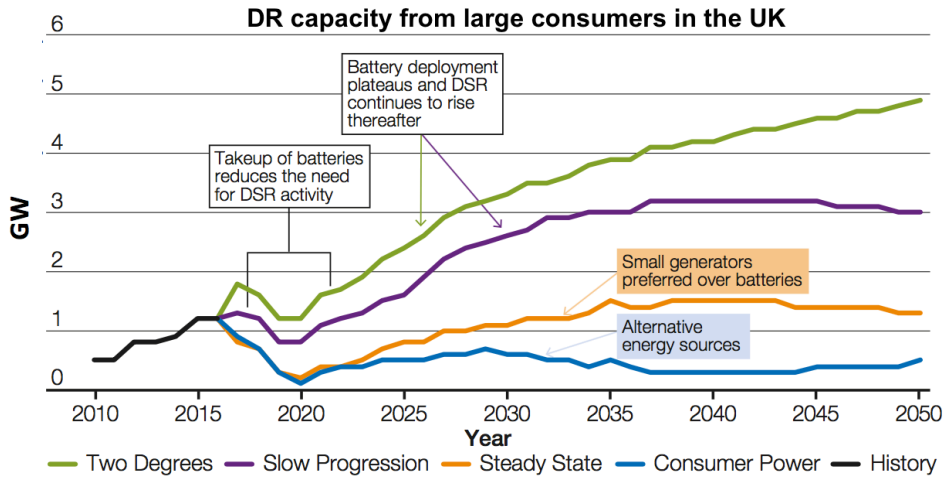

Figure 21: DR capacity from large consumers in the UK projected to different future energy scenarios [30]. Note: DSR refers to *demand-side response,* same as DR.

2.6.2.1 Frequency response programs

Firm Frequency Response (FFR). A monthly electronically tendered service through which NG procures DR capacity that can respond within either 10 or 30 seconds. For more details, see [22].

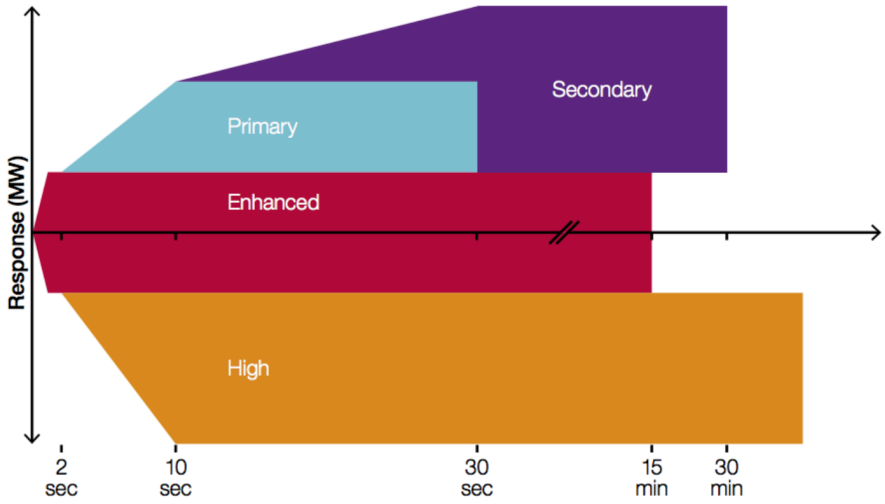

Figure 22: EFR within the existing frequency response services in the UK [16]

Enhanced Frequency Response (EFR): A fast frequency response program, which requires consumer energy assets to provide full response to low and high frequency events in less than a second, sustained for 15 minutes. This is represented within the existing frequency responsive grid balancing services in Figure 22, that includes: 1) primary low frequency response – 10 seconds response to low frequency events sustained for 30 seconds, 2) secondary low frequency response – 30 seconds response to low frequency events sustained for 30 minutes and 3) high frequency response – 10 seconds response to high frequency events sustained indefinitely. For more details, see [33].

Frequency control by demand management (FCDM): This provides frequency response through interruption of consumer load when the system frequency dips below the low frequency relay setting on site. The interruption is for 30 minutes duration, 10 to 30 times per annum.

2.6.2.2 Reserve programs

Short Term Operating Reserve (STOR): This source of reserve energy is procured via three tenders throughout each year where a response time of less than 20 minutes is preferable.

Fast Reserve: A monthly tendered market designed to procure large blocks of reserve power of 50MW to respond within two minutes.

Demand Turn Up: A service which will pay consumers to increase their load when there is excess energy in the system, within several hours of a signal.

2.6.2.3 Capacity market programs

Capacity Mechanism: A catch-all term for the auctions for the capacity market that NG runs to guarantee capacity for any given year.

Transitional Arrangements: Auctions that are in place to help consumers enter the capacity mechanism in the same way as the main auction, but for a much shorter term.

2.6.2.4 Peak management programs

Triad Avoidance: Reducing consumption at periods where peak winter national demand is forecast, in order to proportionally reduce transmission network use of system (TNUoS) charges.

Red Zone Management (RZM): Shifting consumption to avoid periods of highest distribution network cost distribution use of system (DUoS), often referred to as 'red-zones'.

2.6.3 DR programs in mainland Europe

Figure 23 maps the status of DR in the EU. France, Belgium, Switzerland, Finland and Ireland have commercially active DR programs, while there are partial market openings in Germany, Netherlands, Norway, Sweden, Denmark and Austria. DR programs in mainland European countries are operated by the respective TSOs and are commonly designed around ancillary services such as frequency response and reserves. Balancing market participation is encouraged by a few TSOs. Some of the DR programs, participation requirements and incentives in these countries are listed in Appendix A [34].

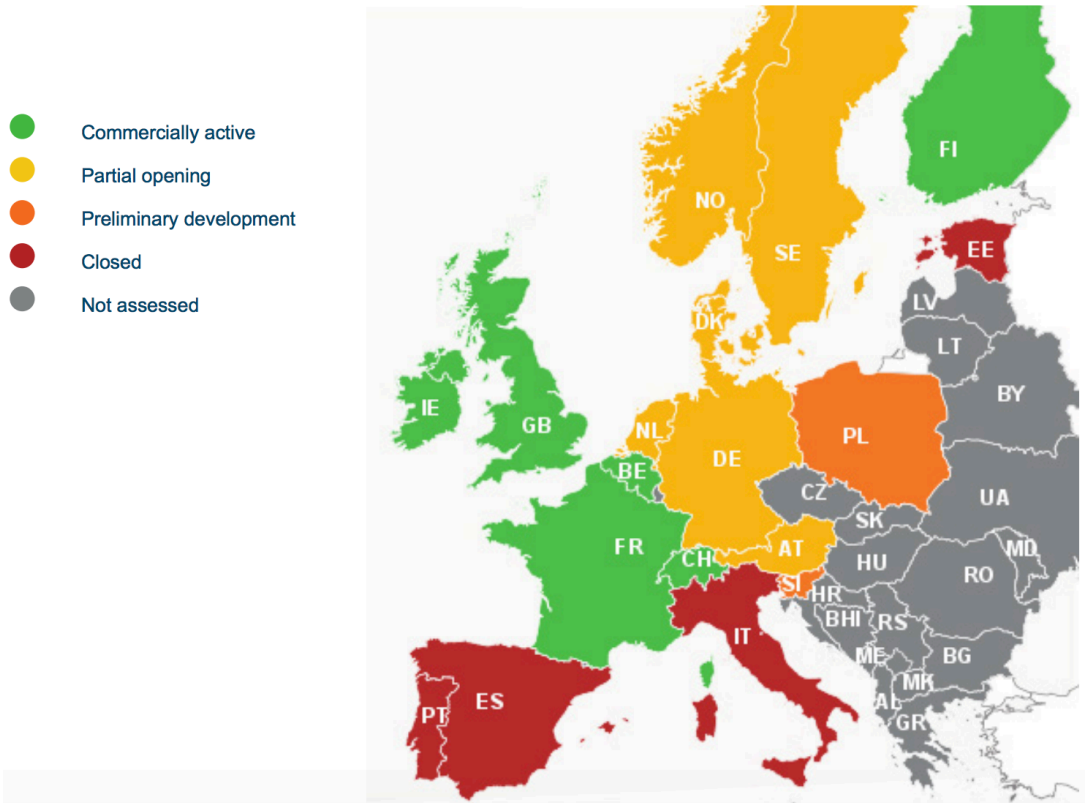

Figure 23: Status of DR programs in EU [34]

Chapter 3

Literature review

3.1 Introduction

This chapter reviews the literature that aligns with the concepts and motivation of the presented research. It aims to provide an overview of the related works and addresses them in the context of the specific thesis objectives. Building load estimation models and their applications are reviewed in Section 3.2. These models are also evaluated for their applicability in DR operations explored in the thesis. The applications of supervised machine learning (ML) algorithms and model calibration methods in building load estimation studies are reviewed in Section 3.3. Data-driven modelling applications in the area of DR have been summarised in Section 3.4. These include studies on DR capacity scheduling and performance evaluation. State-of-charge (SoC) management strategies for battery energy storage systems (BESS) based DR are also reviewed in this section.

3.2 Building load estimation

In this section, the different approaches to building load estimation modelling and their applications are discussed. These approaches are categorised into physics-based (white-box), hybrid (grey-box) and data-driven (black-box) models as shown in Figure 24.

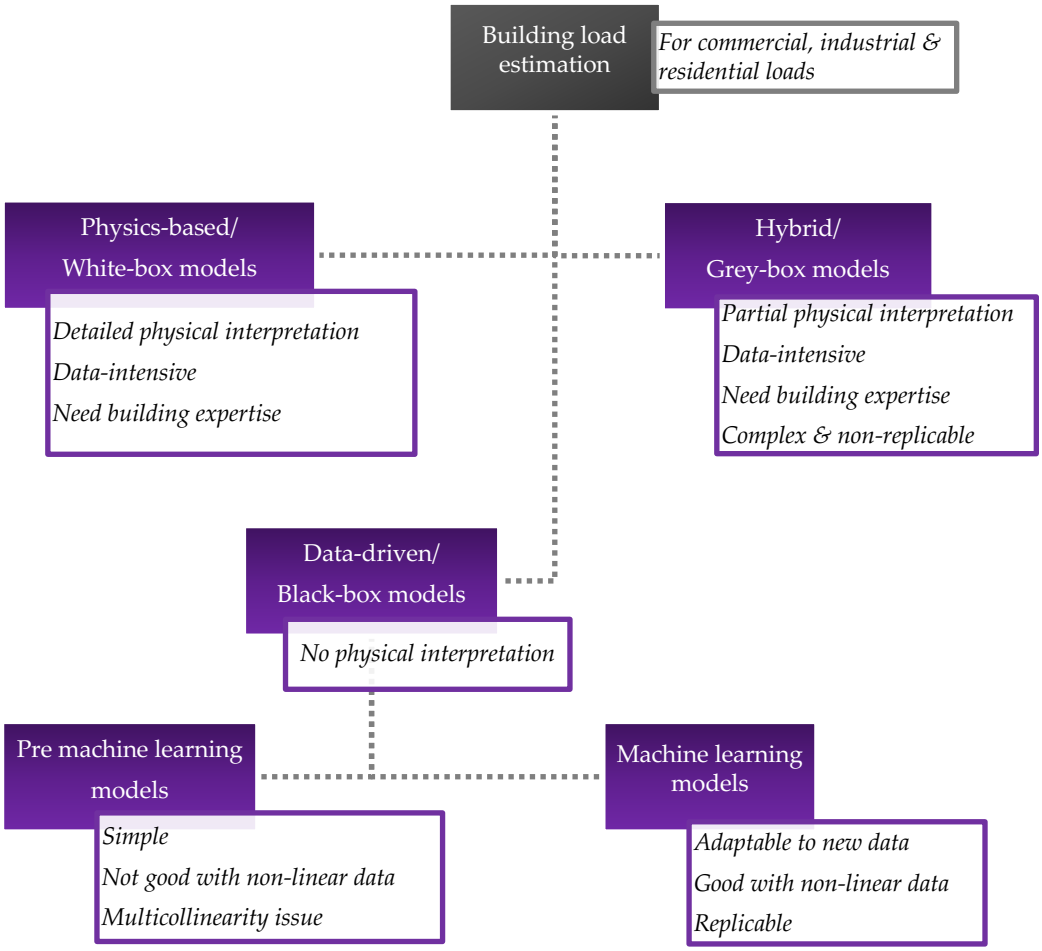

Figure 24: Building load estimation models and their features (*italics* box).

3.2.1 Physics-based models

The generation of physics-based models is the most comprehensive approach towards building simulation for varied applications. Since physical interpretation is possible, these are also referred to as white-box models. They are developed using numerical software that solve equations describing the thermal behaviour of buildings. According to [35], three main methods are used within the physics-based models:

- ***Computational fluid dynamics (CFD) method***: This is recognised as a detailed three-dimensional method based on the decomposition of each building zone to large number of control volumes. The CFD method is particularly applied to describe air/pollutant flow profiles, concentration distribution and temperature distribution in complex building geometries. The complexity of model implementation demands high computational times and knowledge of fluid dynamics. Available software based on this method are FLUENT, COMSOL Multiphysics, MIT-CFD and PHOENICS-CFD.
- ***Zonal method:*** This is a simplification of the CFD method and a faster way to detail the building indoor environment, where each building zone is divided into cells. A portion of a room could be considered as a cell. It is used to evaluate the spatial distribution of temperature, pressure, air velocity, concentration and also visualise indoor air flows. One of the requirements of this method is the previous knowledge of flow profiles, which is not easily available. SimSPARK is an example software based on this method.
- ***Nodal (or multizone) method:*** In this method, each building zone is assumed to be a homogenous volume (or node) of uniform state variables such as temperature, pressure, concentration, etc. A node represents a room, a wall or something more specific like the loads. This is considered as a one-dimensional method and has the least computation time compared to CFD and zonal methods. The nodal method is particularly adapted for estimating building load, energy costs and room temperature, among others. Few example software based on this method are TrnSys (Transient Simulation Program), EnergyPlus, e-Quest, IDA-ICE, ESP-r, Clim2000, BSim and BUILDOPT-VIE.

Among the three physics-based methods discussed above, nodal method is the most preferred for building load estimation in the literature. Very often, an electrical analogy of the physical problem is adopted to simplify the calculations and hence reduce the computation times. Since the applicability of nodal method to large volume spaces is limited, studies based on a combination of CFD with nodal methods could also be found [35].

Building load estimates from physics-based models are highly accurate since many energy processes are taken into consideration during their development. The need for model development from the ground up is minimised since a variety of building simulation software capable of doing sophisticated calculations are available. Generally, the following data are collected before model simulation and subsequent load estimation: weather variables (ambient temperature, humidity), geographic (location, orientation), building design (geometry), thermo-physical variables (based on building materials), characteristics of the HVAC system, occupancy information and operating schedule [7]. Such simulated models do not require measured data as input and hence enable load estimation for buildings that are being planned, constructed or renovated. As a result, they enable designers to make alterations to the buildings based on the desired energy performance. Nevertheless, if the measured data are available, these may be used for model calibration. This helps guarantee that they closely represent the actual building behaviour [36]. Model calibration also ensures compliance with the Measurement and Verification (M&V) guidelines on quantifying the energy and cost savings resulting from improvements in energy consuming systems. The International Performance Measurement and Verification Protocol (IPMVP) and American Society of Heating, Refrigerating and Air-conditioning Engineers (ASHRAE) *Guideline 14* have documented some of these [37].

Physics-based building load estimation models have been developed for diverse applications such as energy performance evaluation, energy cost calculation, selection of appropriate physical components and testing of control techniques, among others. According to [38], the physics-based models developed for individual buildings help understand the characteristics of representative building classes. This bottom-up approach allows further extrapolation to regional or national scales. Many studies using physics-based models, have explored the impact of physical components such as energy efficient walls/roofs, fenestration (arrangement of windows, doors) technologies, thermal insulation materials and others on building loads. In the literature review in [39], simulations of the influence

of phase change materials, based on tools such as EnergyPlus, TrnSys and ESP-r are observed to be useful in analysing the potential of passive cooling/heating towards energy savings. The review in [40] notes that the simulation of different lighting technologies and control techniques helped assess the lighting comfort levels and energy consumption of buildings. In addition to the tools discussed earlier, MATLAB/Simulink, LabVIEW and lighting tailored simulation tools such as Radiance were also used in the reviewed studies. Apart from the physical components of the building envelope, the influence of weather variables on building energy have also been explored using physics-based models [41].

3.2.2 Data-driven models

Data-driven models used in building load estimation do not require any physical information, heat transfer equations or thermal parameters. These are based on the implementations of functions deducted from samples of measured data describing the behaviour of a building load. Due to the lack of physical interpretation, these may also be referred to as black-box models. Data-driven models are classified into pre – machine learning (pre-ML) models and ML models. This classification has been adopted in the thesis to distinctively identify the models based on the period during which they were of interest to the building energy modellers/researchers. There are obvious overlaps in the models and these are clarified in the discussions below.

3.2.2.1 Pre – machine learning (pre-ML) models

Pre-ML data-driven models were based on multiple linear regression (MLR) and timeseries analysis methods. These are discussed further.

3.2.2.1.1 Multiple linear regression (MLR)

The MLR method uses predictor variables such as ambient temperature, humidity, time-of-day, etc. to describe the building load and derive regression equations [42]. An MLR equation estimates the building load $f(x)$ based on a linear combination of p predictor variables and their corresponding regression parameters β_j as:

$$f(x) = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \ldots \ldots + \beta_p x_p \quad (3.1)$$

These parameters are tuned based on the measured building load using an optimisation method and do not have any resemblance to the physical parameters used in physics-based models. Residual (error term) of the estimated building load is the difference between its measured value and estimated value. Residuals can be used to assess the scope for improvement in a regression model. Figure 25 shows the correlation and residual plots for an MLR based building load estimation model using ambient temperature as a predictor variable.

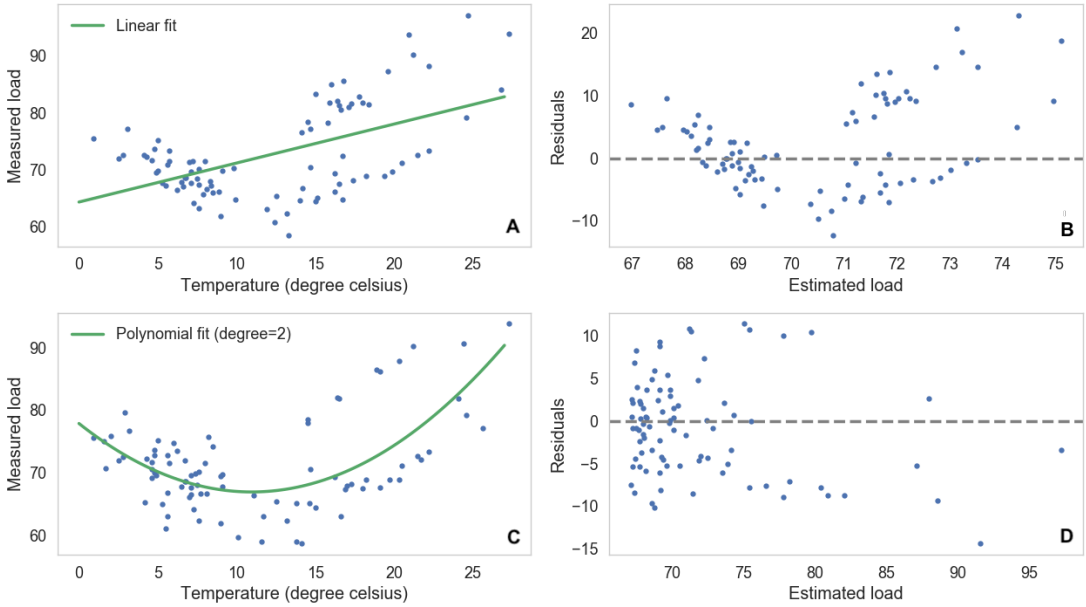

Figure 25: Correlation and residual plots for an MLR based building load estimation model using linear fit (plots A and B respectively) and polynomial fit (plots C and D respectively) with ambient temperature as the predictor variable.

In the above given figure, plot A correlates the building load with the ambient temperature and shows that a linear fit doesn't sufficiently explain their relationship. Residuals based on the linear fit model shown in plot B are unevenly distributed around the residual value of 0 (where estimated value is equal to measured). This suggests that further improvement to the model is necessary to capture the true relationship between the ambient temperature and the building

load. Using a polynomial fit of degree 2 shown in plot C better explains the non-linear relationship between these variables. Residuals based on the polynomial fit model shown in plot D are more evenly distributed than those based on the linear fit model. This example highlights that, transformations (such as using polynomials) of variables are necessary to capture non-linearity and improve the MLR model. While developing building load estimation models using MLR, this issue should be addressed on an individual basis for each building data. Because of this, MLR based models are difficult to be replicated on large number of buildings.

In addition to this, the issue of multicollinearity (one variable influencing the other variable) also needs to be addressed before deriving regression equations [35,43]. Nevertheless, MLR has been employed in many building energy studies due to its ease of interpretation and computational simplicity. Since MLR is considered as a simple regression algorithm in the context of ML, the applications are discussed along with other algorithms in Section 3.3.

3.2.2.1.2 Timeseries analysis

Timeseries data are ordered sequences of values recorded at equal intervals of time. Building electricity meter data is an example. Timeseries analysis methods decompose the timeseries data into trends (linear increase or decrease), seasonality (variations over day or seasons) and residuals (unexplainable part of the timeseries). Autoregressive integrated moving average (ARIMA) is the most common approach. ARIMA is based on the idea of removing non-stationarities (i.e. mean, variance, etc. not constant over time) from the timeseries by a differencing process. It is complex and time-consuming. Many variants of the ARIMA model (ARIMAX, SARIMA, etc.) have limited structural interpretation and performance of these models can be affected by the outliers in the data [44].

3.2.2.2 Machine learning (ML) models

ML is an interdisciplinary field which borrows ideas from statistics, computer science, engineering, cognitive science, optimisation theory and many other disciplines. With progress in smart grids, availability of smart meter data and

development of competent ML algorithms, pre-ML models discussed above have given way to ML based data-driven modelling. The basic concepts behind MLR and timeseries analysis are valid in the context of ML, i.e. inferences are made from samples of training data based on the influence of other variables. However, instead of the static data considered in the pre-ML models, ML is oriented towards adaptation to or learning from new data. Further in this thesis, any references to data-driven models are understood to be those based on ML algorithms.

ML algorithms are broadly categorised into supervised and unsupervised [45].

Supervised machine learning: In supervised ML, the algorithm learns through training data and the goal is to learn to predict a desired outcome based on new data. If the algorithm is predicting a numeric outcome or a continuous response, it is referred to as supervised *regression*. On the other hand, if the algorithm is predicting a categorical response, it is referred to as a supervised *classification*. These are illustrated in Figure 26.

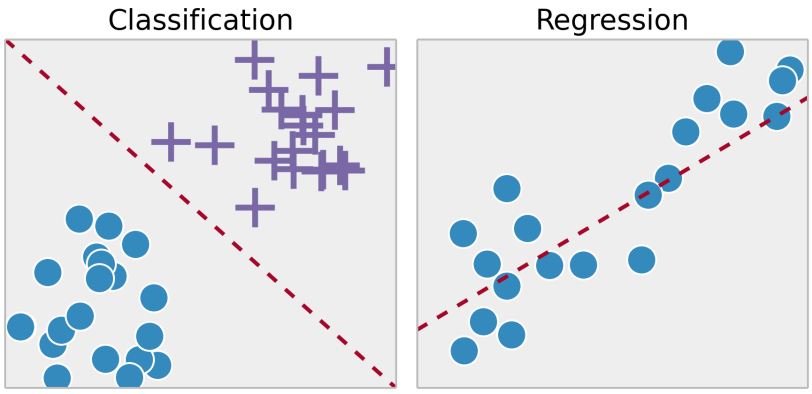

Figure 26: Supervised classification and regression on a two-dimensional plot [46]

Unsupervised machine learning: In unsupervised ML, the algorithm infers certain information from the input data that helps in decision making, predicting future inputs and efficiently communicating. Two classic examples of unsupervised ML are: 1) *dimensionality reduction*, where high dimensional data with large number of

variables are reduced to few principle variables and 2) *clustering*, where the goal is to discover similar examples from the input data.

In the literature, ML based data-driven building energy models have been developed for a variety of applications such as load estimation, energy pattern profiling, regional energy consumption mapping, benchmarking and development of retrofit strategies. The scope of these studies ranges from sub-systems in individual buildings to large number of buildings in the macro-level. Supervised ML algorithms of interest include artificial neural networks, support vector machines and decision trees, among others. Unsupervised ML based dimensionality reduction methods such as principal component analysis and clustering methods such as self-organising maps, K-means and hierarchical clustering have also been implemented in some of these studies [47]. Since the presented research focuses on supervised ML algorithms, a detailed review of their applications is given in Section 3.3.

3.2.3 Hybrid models

Hybrid or grey-box models are developed based on the coupling of physics-based and data-driven models. These are particularly useful when sufficient thermo-physical data are not available, as in the case of existing buildings. In such instances, some of the building characteristics are parametrised using techniques such as optimisation. Knowledge of the boundary values of parameters eliminates the need to input precise values during simulation. The hybrid models are better at associating thermo-physical data with end-uses. These are also applicable for monitoring and control of the building physical components. Hence, hybrid models overcome some of the limitations of the physics-based models. Although not as descriptive as physics-based models, these allow physical interpretation through a rough description of the building geometry and thermal behaviour [35].

Plenty of hybrid models based on coupling of physics-based models with optimisation techniques were developed in the literature. In many of these studies,

the optimisation technique contributed to parameter estimation for the physics-based model based on objectives such as minimising energy consumption, minimising costs or maximising thermal comfort, among others. For example, given a target energy consumption level, the optimal thermal properties (such as conductivity) of the building walls could be calculated. In certain other studies, a physics-based model initially generated a set of data that were used as input parameters to derive regression equations. This helped minimise the need to run the physics-based model repeatedly and also reduced the computation times significantly [35]. These capabilities of hybrid models make them good candidates for calibration, monitoring and control applications.

3.2.4 Evaluation of building load estimation models

Reliability, adaptability, computational efficiency, cost effectiveness, potential for large scale deployment and automation are some of the criteria for selection of building load estimation modelling performed in this research. The physics-based, hybrid and data-driven models are evaluated based on these criteria towards application in DR operations addressed in the thesis.

Physics-based models demand inputs such as weather variables, geometrical data, thermo-physical variables, occupancy status, load characteristics, among others to provide very accurate building load estimates. The level of building information required for developing reliable physics-based models is not always available nor accessible from large building consumers. This may be due to the total absence of such information or due to privacy concerns and the reluctance to share data with third parties such as DR aggregators. With increased deployment of sensing devices and improved privacy measures, this may change in the future. However, as of today, the lack of required building information leads the building energy modeller to make many assumptions. The uncertainties induced by these assumptions have negative consequences on the model performance. Hybrid models may help improve these assumptions by generating boundary values using optimisation techniques. However, such approaches are unique to each building and cannot be

easily replicated on large number of buildings as required in DR. With time, building energy consumption may change (due to addition of a new load or retrofitting) and the models deployed in DR operations will need to adapt to such changes. Physics-based and hybrid models are not easily adaptable and reasonable expertise in building thermo-physical systems is required to perform model calibration. This inhibits the potential for automation in applications such as DR where the participation of building consumers is ever increasing.

Against this, data-driven models based on ML are interesting for building load estimation in the DR context. As demonstrated in this research, supervised ML models can use predictor variables such as time-of-day and weather to deliver reliable building load estimation. Due to the minimal requirement for building information (except for the smart meter data) customised models can be easily developed. Further, computational complexity of data-driven modelling can be adjusted according to the resources available. This is highlighted in the presented research based on the use of a random-search hyperparameter optimisation method for model calibration. Due to the minimal building information requirement, computational efficiency and cost effectiveness, data-driven modelling can be replicated on a large number of buildings, making it appealing for many DR applications. Data-driven models continuously learn from the new incoming data (such as smart meter data) and are adaptable to changes in building energy consumption even without the need for physical inspection or interpretation. This makes it interesting for deployment in DR operations. When structural changes occur to the incoming data, decline in predictive performance may occur. This can be tracked with reference to a benchmark performance and the model can be calibrated for redeployment, increasing the scope for complete automation.

Data-driven modelling meets most of the criteria for application in DR operations addressed in the thesis and are hence prioritised over physics-based and hybrid modelling for building load estimation. This does not mean that data-driven models are universally applicable for all the DR operations.

3.3 Data-driven modelling for building load estimation

The thesis focuses on the application of supervised ML algorithms for building load estimation. A high-level description of some of the important supervised ML algorithms are given in Section 3.3.1. Studies on their applications in building load estimation are reviewed in Section 3.3.2. Since the presented research explores a computationally efficient data-driven modelling method, model calibration methods are separately reviewed in Section 3.3.3

3.3.1 Supervised ML algorithms

A general description of supervised ML can be summarised as follows. Given a D dimensional dataset (i.e. D number of predictor variables or features) consisting of N data samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ and corresponding response values $y_1, y_2 \ldots . y_N$ in the training set, a supervised ML algorithm finds a functional mapping $f(\mathbf{x})$ that minimises the prediction error through optimisation of a loss function. Although not an exhaustive list, some of the supervised ML algorithms applied in building load estimation studies are introduced below.

Multiple **linear regression** (MLR) discussed in the context of pre-ML data-driven modelling (Section 3.2.2.1.1) is based on the ordinary least squares method in which the sum of squared errors is used as the loss function. Linear regression algorithms such as lasso and ridge are based on penalising methods that add a regularisation term to this loss function. Lasso regression performs L1 regularisation that pushes the small weights of certain features to zero and hence enables feature selection. Ridge regression performs L2 regularisation that penalises large weight values. This is useful when there is collinearity between the selected features (multicollinearity) [48]. Elastic-net regression performs L1 and L2 regularisation, hence combining their benefits such as embedded feature selection and robustness to multicollinearity [49]. However, linear regression algorithms are applicable to non-linear data only with appropriate transformation of the variables, as discussed previously in Section 3.2.2.1.1.

An **artificial neural network** (ANN) consists of input, hidden and output layers with processing elements called neurons, as shown in Figure 27. Multilayer perceptron (MLP) is the most popular architecture in ANN. Other architectures of interest include Bayesian neural network (BNN), radial basis function neural network (RBF) and generalised regression neural network (GRNN). ANN is largely preferred due to its suitability for non-linear data, however at the cost of reduced interpretability [50]. The ANN architecture based on MLP is discussed elaborately in Chapter 4 Section 4.3.2.1.

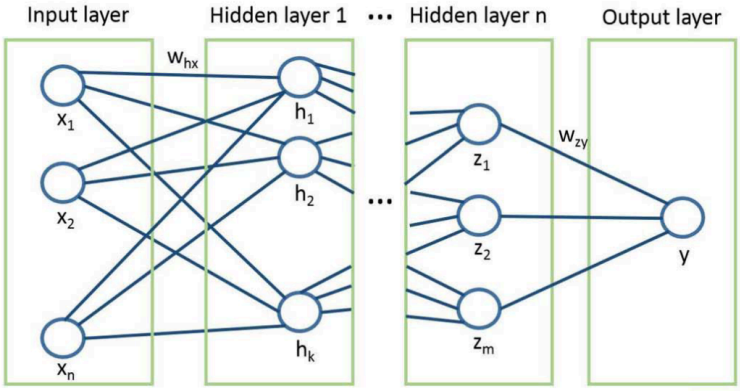

Figure 27: Artificial neural network with two hidden layers [51]

Support Vector Machines (SVM) algorithm is based on the use of a high-dimensional feature space formed by transforming the original predictor variables, and penalising the resulting complexity using a regularisation term added to the error function. SVM is good with non-linear data and also works well on a smaller number of training samples [50]. The SVM architecture is discussed in detail in Chapter 4 Section 4.3.2.2.

Decision trees algorithm, starting from a root node, generates a set of if-then-else rules (represented in Figure 28) using predictor variables at each decision node of the tree such that the response variable data are split and allocated into new nodes. The algorithm pursues a recursive process of data splitting based on a criterion such as mean squared error until it fails at an endpoint called leaf node. The depth of the

tree increases the complexity of decision rules [52,53]. The decision tree algorithm is explored in depth in Chapter 4 Section 4.3.2.3.

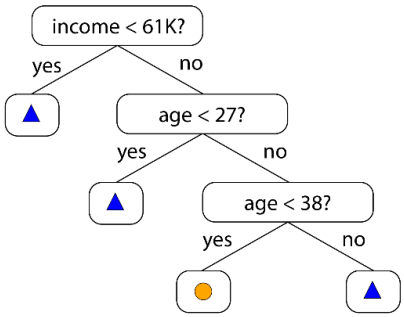

Figure 28: If-then-else logic rules used in decision trees algorithm [54]

Gaussian processes (GP) regression is a fully probabilistic Bayesian model. This means that, instead of finding a single estimate of the functional mapping $f(\mathbf{x})$, GP finds a probability distribution $p(f)$ over likely functions. This is in contrast to most other regression algorithms that find a single best estimate of $f(\mathbf{x})$. In GP, the functional mapping estimates form a multivariate Gaussian distribution $N(\boldsymbol{\mu}, \boldsymbol{K})$ with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{K}$. The covariance matrix is constructed from a covariance function, the selection of which is an important aspect of GP based model development [55]. Since GP prediction is probabilistic, empirical confidence intervals can be computed and the model can be refit in the region of interest. However, some of the downsides are that: 1) the algorithm loses efficiency with high-dimensional datasets (i.e. large number of features), and 2) the algorithm is not sparse, meaning that it uses all training data samples to perform the prediction, making it computationally demanding [49].

Nearest neighbours algorithm finds a predefined number of training samples closest in distance to a query point and bases its prediction on these. Standard Euclidean distance is the most common choice for measuring the distance. The number of samples can be a pre-defined constant or can vary based on the local density of points (such as based on radius). The computational load in finding the

neighbours and storing the entire training set is usually considered as a drawback of the nearest neighbours algorithm. For a dataset with *N* samples and *D* predictors, the algorithm requires *ND* operations to find the neighbours per query point [56].

Models based on **deep learning** adopt the deep architecture for knowledge discovery, wherein the input data is transformed multiple times, while most other algorithms pursue one or two rounds of input data transformation (shallow). Deep learning neural network (DNN) for regression is comparable to the ANN, but with more complex architectures and training schemes, hence enabling multiple levels of abstraction. Unlike ANN, DNN models may have many hidden layers and each layer may have different functions [57].

While most of the ML algorithms are single algorithms, **ensembles** use one or more single algorithms to build base models for learning**.** The gradient boosted trees (GBT) is such an ensemble algorithm based on decision trees base models [50]. These are elaborated in Chapter 4 Section 4.3.2.3.

3.3.2 Applications of supervised ML for building load estimation

Data-driven building load estimation studies have been extensively reviewed in [42–44,47,58,59]. Figure 29 shows the summary of ML based building energy models from [47]. Majority of the models were based on large consumer (non-residential) buildings (81%). Whole building load models were larger in number compared to those based on category loads such as heating, cooling and lighting. A large number of models have used real data instead of simulated or public benchmark data, and the most available data resolution was hourly.

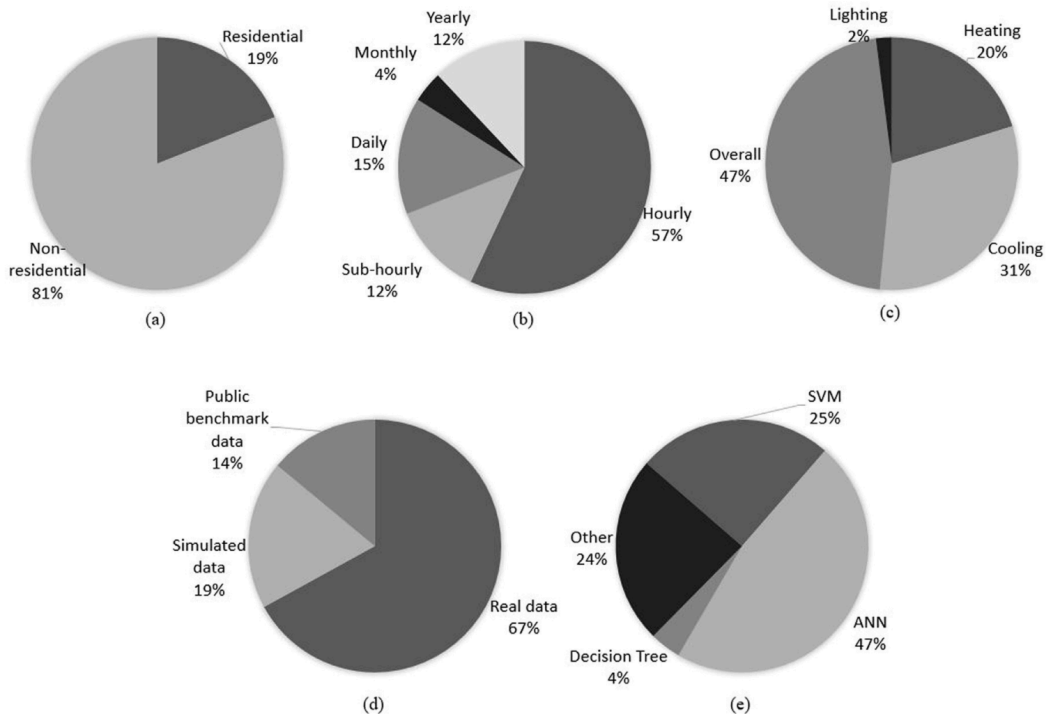

Figure 29: Data-driven building energy models in literature categorised according to (a) type of building, (b) data resolution, (c) type of energy consumption, (d) type of data, (e) machine learning algorithm [47]

In the literature, some of the ML based data-driven models have been developed for specific application areas such as performance measurement and verification (M&V), building control and demand-side management, among others. However, a significantly large number of the models focused on demonstrating the capability of ML algorithms and techniques in performing building load estimation rather than focusing on specific applications. In many of the studies, the model performances were compared and contrasted with the physics-based, hybrid, pre-ML or other ML based models.

Some of the application specific models in the literature are reviewed below.

Data-driven building load estimation models have been applied for **performance measurement and verification (M&V)** practices in the context of IPMVP and ASHRAE Guideline 14 (previously mentioned in Section 3.2.1, for more details refer [37]). In this application, the estimated building baseline loads are used to verify the

impact of efficiency measures on energy and cost savings. Dong et al [60] examined the applicability of SVM algorithm with radial basis function kernel in building baseline load estimation for M&V applications on four commercial buildings in Singapore. Predictor variables such as monthly mean outdoor dry-bulb temperature, relative humidity and global solar radiation were employed in the models, which showed reliable predictive performances. According to the authors, the smaller number of hyperparameters for SVM makes the model calibration simpler in comparison to ANN and genetic programming. Heo and Zavala [61] developed a Gaussian process (GP) based data-driven model to determine energy savings and uncertainty levels in M&V practices. Predictor variables such as ambient temperature, relative humidity, occupancy count and supply air temperature were used at hourly resolution to estimate the building loads. The GP models based on Bayesian principles were demonstrated to capture complex non-linear and multivariable interactions of the building energy behaviour and perform better than linear regression models. The authors claim that GP models can ultimately lead to significantly less expensive M&V practices, but also admit that the results are sensitive to input data quality. Following this, Burkhart et al [62] presented a Monte Carlo expectation maximisation framework for constructing baseline Gaussian process (GP) models under uncertain/sparse input data and demonstrated that it yields more reliable predictions and confidence levels in comparison to the standard GP models as developed in [61].

Building control based on data-driven models is a growing area of application in the literature. Peng et al [63] proposed a data-driven control strategy to increase the efficiency of HVAC systems by accommodating occupants' behaviour in real time. The occupancy-related information learned by supervised and unsupervised ML algorithms is used to infer real-time room set-points for controlling the space cooling system. This data-driven strategy is reported to achieve between 7-52 % energy savings as compared to scheduled cooling operations. Drgona et al [64] explored the use of supervised regression models based on deep time delay neural networks (TDNN) and decision trees to mimic the complex behaviour of model

predictive controller (MPC) for building control. Feature selection was performed based on prior engineering knowledge, principal component analysis and dynamic analysis of the building model. Based on simulations, TDNN-based controllers were able to maintain high comfort and energy savings with a small loss of performance compared to the original MPC and the decision trees model. Guo et al [65] developed building heating load estimation models using supervised ML algorithms such as extreme learning machine (ELM), MLR, SVM and ANN. Predictor variables such as weather, operational schedules and indoor temperature were selected based on correlation analysis. The study proposed a strategy for obtaining the thermal response time of the building, which was estimated to be 40 minutes, and used as forecast horizon for the data-driven model. Among the different algorithms, the performances of ELM models were found to be superior.

Many data-driven building load estimation studies have highlighted **demand-side management (DSM)** as a potential application area. Chae et al [66] developed a day-ahead building load estimation model at 15 minutes resolution using predictor variables such as day type indicator, time-of-day, HVAC set temperature schedule, outdoor air dry-bulb temperature and outdoor humidity, based on the ANN algorithm. The proposed model can predict daily load profiles as well as peak electricity consumption reasonably well for an office building and can be used for decision making in contexts where energy bills are determined based on peak consumption at sub-hourly intervals. Jetcheva et al [67] addressed the need for demand-side flexibility and developed ANN based ensemble models with ambient temperature inputs for day-ahead load estimation in six large consumer buildings and demonstrated that they outperform the timeseries SARIMA models. Leung et al [68] developed an ANN Levenberg–Marquardt backpropagation algorithm based hourly and daily cooling load estimation model for a university building. External weather variables such as dry-bulb temperature, wet-bulb temperature, global solar radiation, rainfall, clearness of sky, cloud condition and wind speed and internal occupancy space load data were used to develop reliable models. The occupancy data was derived from a building power monitoring system, part of the advanced

metering infrastructure in the building, and helps improve the model's predictive performance towards better DSM practises.

ML based modelling has also been used to study the **impact of climate change** on building energy use. Lam et al [69] investigated how the energy use in an office building in Hong Kong, based on simulation (using VisualDOE 4.1 tool) can be correlated with a new composite climatic index Z derived using principal component analysis on major climatic variables such as temperature, humidity and solar radiation. Multi-year (1979–2007) building energy consumption were correlated with Z using linear regression models and showed an increasing trend indicating a gradual change of climatic conditions that might affect energy use in buildings in the future.

While there exist other application areas for data-driven building load estimation, a significant number of models were developed around specific ML algorithm/s without any particular application focus as mentioned earlier. These models may however be suitable for different applications depending on the modelling methodologies adopted. Some of these studies are reviewed below.

Starting in the early 1990's, ANN has been the most widely implemented ML algorithm in building load estimation studies due to its applicability on non-linear data and reliable predictive performance. The SVM regression algorithm, highly applicable on non-linear problems even with small quantities of training data has been used for building load estimation since mid 2000's. The numerous studies reviewed in [70,71] have shown superior performances for the ANN and SVM based models individually and in comparison to physics-based and pre-ML models. A variety of large consumer buildings were considered and predictor variables such as local weather and building conditions were usually employed in these studies. Decision trees based building load estimation studies [72–74] preferred this ML algorithm for its logic based structure that improved interpretability and computational efficiency. Some of these models were also deployed in building energy management systems. Application of Gaussian process algorithm in large

consumer buildings demonstrated significant load estimation capabilities for different forecast horizons [75,76]. A nearest neighbours based model was developed [77] for day-ahead load estimation on a variety of buildings using historic meter data as the only predictor variable with reliable performance. Studies [78,79] preferred extreme learning machine algorithm over ANNs for its faster learning rate and better predictive performances for building data analysed. A deep neural network (DNN) for supervised regression and unsupervised feature selection was implemented on a building load estimation study [80]. Simpler ANN with two hidden layers outperformed the performance of the DNN model, highlighting that deep models are not necessary while analysing one year long building datasets.

There is rather no universally best ML algorithm. A data-driven model's forecast performance on a given building dataset primarily depends not only on the ML algorithm, but also on the hyperparameter values, the predictor variables and the sizes of train-test sets, among others. Taking these factors into consideration, the presented research develops a data-driven building load estimation model for large consumers with specific application in DR operations.

3.3.3 Model calibration methods

An ML algorithm learns a specific set of *parameters* such as weights or coefficients from the dataset. In addition to these parameters, the ML algorithm also has controls called *hyperparameters* that determine how it learns and forecasts [81]. It is known that the optimal values for hyperparameters give an ML algorithm the highest predictive performance on a given dataset and bad choice of these values would result in poor models. The problem of identifying optimal values for hyperparameters of an ML algorithm on a given dataset is called hyperparameter optimisation (HPO) [82]. An ML model with optimised hyperparameter values is referred to as a *calibrated* model in this context.

The data-driven building load estimation modelling in this research provides a framework where multiple ML algorithms could be compared and contrasted. When comparing the performance of different ML algorithms on a given dataset, it is important to make the comparison across their *calibrated* models. It is observed that such attempts have not been made in previous studies [14,25], pointing at a gap between empirical and applied ML. Nevertheless, calibration of ML models based on HPO methods has been adopted in building load estimation studies. The review presented here attempts to compare and contrast the HPO methods used in data-driven building load estimation studies (summarised in Figure 30)

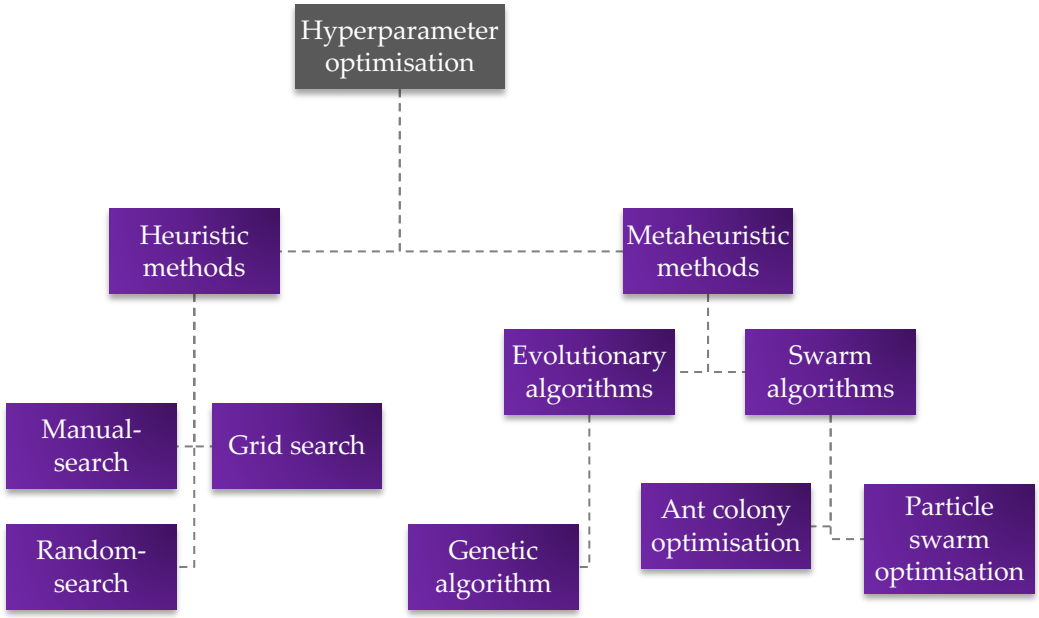

Figure 30: Prominent hyperparameter optimisation methods used for calibration of data-driven models in the literature

The HPO methods found in load estimation literature could be categorised into heuristic and metaheuristic methods. The heuristic method of finding an optimal value is problem-dependant and focuses on human intuition, insights and learning. In contrast, a metaheuristic (*meta* in Greek means *beyond*) method provides a problem-independent framework or a set of guidelines to develop implementations

of heuristic optimisation algorithms. The evolution of these optimisation terminologies and their distinctions are discussed by Sörensen [85].

In either of the heuristic or metaheuristic methods, the optimal hyperparameter values are found from the pool of predetermined trial values. The most primitive form of heuristic HPO method is manual-search. Due to the simplicity of this method, it is widely adopted in building load forecast studies. When there is more than one hyperparameter to be optimised, predetermined trial values of one is combined with fixed values of others to form a search space of hyperparameter sets. These hyperparameter sets are used to train and test the learning algorithm on a portion of the dataset through cross-validation. This is performed by splitting the dataset into many folds, where some are used for training and the rest for testing. From the hyperparameter set delivering the highest forecast performance through cross-validation, near-optimal value for the first hyperparameter is identified. This hyperparameter value is fixed for succeeding rounds where other hyperparameters are optimised the same way. This method is also referred to as stepwise-search.

Using manual-search, Ruiz et al [86] and Li et al [87] found near-optimal values for the number of hidden layer neurons of the ANN learning algorithm on building datasets by keeping other hyperparameter values constant. Dong et al [60] optimised a key hyperparameter for SVM learning algorithm on multiple building datasets and identified different near-optimal values for different datasets. This method was also adopted in other load forecast studies [88,89]. Manual-search enables simple computation, but at the cost of a limited search space. Due to the small search space, only good hyperparameter values that do not render finely calibrated forecast models are achieved. Manual-search makes it difficult to reproduce results since modellers with different expertise may use different sets of hyperparameters in the search space. Manually optimising large number of hyperparameters and trial values that constitute a high dimensional search space is indeed a difficult task [81].

In preference to manual-search, grid-search is a widely adopted heuristic HPO method in which the search space is a grid formed by assembling all combinations (hyperparameter sets) of predetermined trial values [90]. Chae et al [83] optimised an ANN learning algorithm for an office building dataset over the number of hidden layer neurons and other hyperparameters by forming an exhaustive search space. Massana et al [91] used grid-search for ANN and SVM learning algorithms on a university building dataset to calibrate them on datasets formed by different predictor variables. In another building energy study by Rishee et al [92], key hyperparameters of SVM were optimised by iterations through finer grids until an expected threshold of forecast performance was achieved by the learning algorithm. This approach to HPO was also adopted by a building load predictive model based on SVM and random forest algorithm [93]. Grid-search is more reliable and replicable in comparison to a simple manual-search since it cross-validates all possible combinations of trial values. However, the search space grows exponentially with the number of hyperparameters and this curse of dimensionality makes grid-search a computationally slow method [81].

As an improvement over manual-search and grid-search, Bergstra and Bengio [81] developed a heuristic method of HPO called random-search. In terms of performance on high dimensional hyperparameters, random-search HPO was proven to be superior to grid-search based on its application on different ML algorithms and datasets. Unlike grid-search, all combinations of hyperparameter values in the search space are not tried out. Rather a fixed number of sets are randomly sampled, hence making it computationally efficient. The sampling size could be altered based on the available computational resources, giving more control to the modeller.

The metaheuristic methods of HPO implemented in load forecast studies include evolutionary/genetic algorithm [94,95], swarm algorithms [39–43], among others. Liu et al [98] used a swarm algorithm for HPO and recommended to implement the model calibration on a high performance computer due to the complexity in

calculation and bigger storage requirement. A study of different HPO methods on an SVM based energy forecast model showed that metaheuristic methods are computationally slow [101]. This might be because of the exhaustive search performed to find the optimal hyperparameter values. The implementation of metaheuristic methods requires more code complexity, client-server architectures in which a master process keeps track of the trials, a shared database and inter-process communication mechanisms, leading to significant technical hurdles for software engineers [81].

Based on the presented review, random-search HPO is found to be suitable for model calibration in this research.

3.4 Demand response (DR) modelling

In the extensively surveyed literature in [9,10], DR is usually formulated as optimisation problems dealing with objectives such as social welfare (utility) maximisation, cost minimisation, energy consumption minimisation or combinations of these. The studies have also addressed issues such as the impact of price, renewables, storage and electric vehicles in the DR paradigm. Data-driven modelling studies for DR applications are limited in the literature.

The presented thesis applies the developed building load estimation modelling approach in DR operational tasks such as capacity estimation, performance evaluation and reliable operation. In incentive-based DR programs, capacity estimation is the process of estimating the available DR capacity from a consumer based on their energy assets such as flexible loads, standby generators and storage systems. In this context, performance evaluation estimates a baseline load and compares the actual DR delivered with this to evaluate the performance of a consumer. Financial incentives for the DR participant consumers are assessed based on their DR capacity availability and actual delivery. Studies focussed on data-driven DR capacity estimation and performance evaluation were found to be

limited in the literature. The available studies are reviewed in Section 3.4.1. Reliable operation of a consumer energy asset such as battery energy storage system (BESS) in DR can be ensured by proper state-of-charge (SoC) management. The state-of-the-art in SoC management is reviewed in Section 3.4.2 and the lack of data-driven approaches is highlighted.

3.4.1 DR capacity estimation and performance evaluation

Nghiem and Jones [102] developed a Gaussian processes (GP) based data-driven model for predicting the DR behaviour of commercial buildings with a quantifiable confidence measure. The GP regression model used experimental DR signals and weather variables as predictors and building load as the response variable. For a DR aggregator, using this black-box model, it is not necessary to understand the internal DR strategies adopted by each respective building so as to estimate their load response. This approach could be easily integrated into the existing building energy management systems.

Jung et al. [103] estimated the available flexible DR capacity in two large buildings based on a set of rules learnt using a data-driven approach. The data were sourced specifically from temperature/humidity/light sensors, carbon dioxide sensors, passive infrared sensors and smart plug power meters. This method was found to be better than the conventional manual audit processes used to estimate DR capacity.

Behl et al. [104] demonstrated the superior performance of decision trees based data-driven model for baseline estimation on 8 large buildings. The approach was adapted to help choose the best DR strategy based on the state of the building and weather forecasts. The tree-based regression was also used for DR control synthesis in real-time, enabling the trade-off between thermal comfort and load curtailment. This was shown to outperform the rule-based DR control strategy by achieving 17% more load curtailment without affecting thermal comfort.

Chen et al. [105] proposed an SVM regression based data-driven model for baseline load estimation and demonstrated this on four office buildings. The baseline is used to evaluate the DR performance for building load curtailment. The approach was proved to perform significantly better than the conventional baseline models that were based on simple averaging or polynomial regression. The possibility of integrating the SVM model as an online tool within the building energy management system was discussed.

Park et al. [106] showed that a data-driven model based on unsupervised ML techniques has lower errors in comparison to conventional day-matching methods for baseline load estimation on residential buildings. The self-organising map and K-means clustering methods efficiently reduce the large dataset into representative weight vectors and cluster them to match the load pattern with that of the DR event day.

Data-driven capacity estimation studies have seldom focused on computational efficiency and deployment in large scale which is necessary in the context of DR. The presented research attempts to fill that gap. Unlike the data-driven DR performance evaluation studies in the literature, model development methodologies in the presented research are guided by the ultimate need for deployment. Further, the applicability of the data-driven models in the ongoing DR programs in different electricity markets is also explored.

3.4.2 Reliable operation of battery energy storage system in DR

A BESS may be connected to the grid as a standalone unit, collocated with renewable generation plants, or based at consumer premises (also referred to as behind-the-meter BESS). Due to the instantaneous response capability of BESS, these are capable of participation in fast DR such as frequency response. The decline in BESS prices, is in favour of large investments in consumer BESS. Based on a 2017 survey in the UK electricity market, the investment interest in consumer BESS was

the highest (49%) compared to standalone and collocated projects. The survey results are shown in Figure 31.

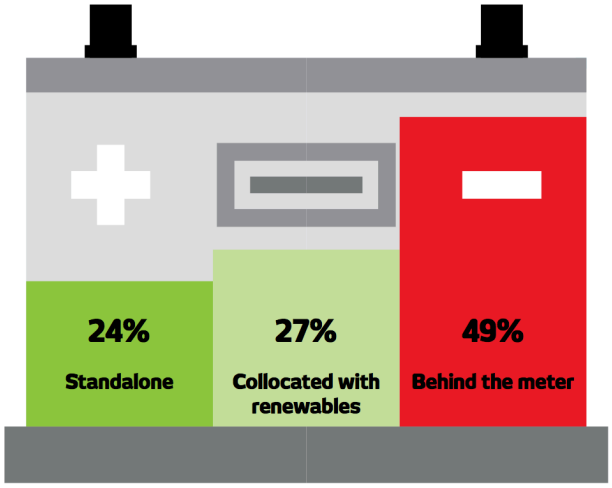

Figure 31: A 2017 survey in the UK electricity market showing the investment priorities in BESS projects [107]

One of the major constraints for BESS in providing DR is the limited energy capacity. State-of-charge (SoC) is the energy remaining in a BESS. In order to sustain a DR service such as frequency response, the BESS needs to adopt certain SoC management strategies.

For better comprehension, the different SoC management strategies adopted in the literature have been categorised into deadband strategy, real-time offset strategy, scheduling strategy and electric vehicle (EV) strategy. These are summarised in Figure 32 and reviewed further.

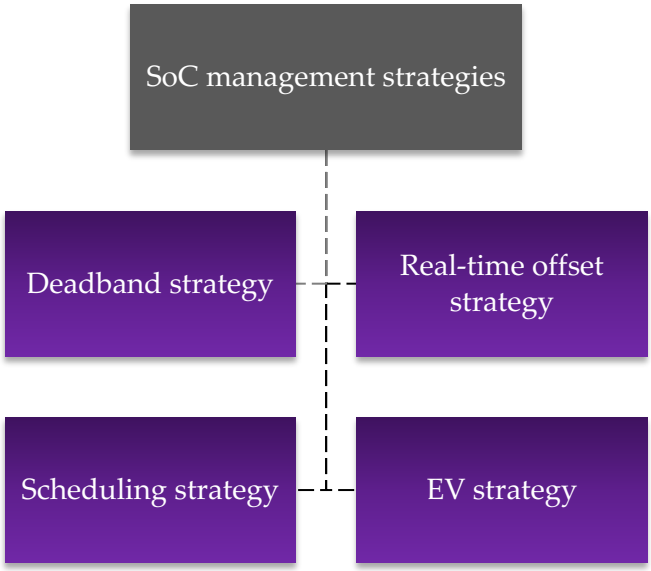

Figure 32: SoC management strategies for BESS based DR considered in previous studies

3.4.2.1 Deadband strategy

Oudalov et al [108] analysed different SoC management strategies for a lead-acid BESS with 70% round-trip efficiency performing frequency response with the objective of cost minimisation to the battery owner in the European market. A no SoC management strategy resulted in the SoC declining in less than a month. Hence the BESS is recharged using a bias power when frequency is within a non-critical frequency window or deadband of $\pm 20\ mHz$ around the nominal frequency. If the SoC exceeds a set-point, the excess energy is sold in the intraday market. Although a substantial waste of energy, emergency resistors were proposed to absorb energy from the grid when battery SoC was 100%. The strategy helped maintain the SoC within a safe range for one month.

Thorbergsson et al [109] adopted an SoC management strategy for frequency response from a Li BESS in the Danish market. The BESS is recharged/discharged when frequency is within the deadband so as to achieve the set-point SoC, although the bias power used is unclear. An alternative strategy of achieving set-point SoC after every service delivery failure was explored. In comparison, the deadband

strategy proved to have lower battery degradation, longer lifetime and higher revenue.

3.4.2.2 **Real-time offset strategy**

Borsche et al [110] proposed the use of an offset power with slower ramp rate on top of the expected response such that it does not interfere with the frequency response service. The offset power is estimated with a delay as the average of the expected response over a given period of time. The delay corresponds to that of the succeeding tier of services such as balancing, trading, etc. that may account for the offset. The averaging forces the frequency signals to be zero-mean and the averaging period influences the ramp rate of the offset. The impact of this SoC management strategy on grid stability was further explored in [111] and found to be negligible.

In [112], California independent system operator (CAISO) considered SoC management for a battery providing frequency response using intraday market based offset power. The energy required to achieve a target SoC was estimated at the market closure time and expected to be consumed from the beginning of delivery until the end of the market interval duration. A similar strategy was adopted by Lian et al [113] for a battery providing frequency response in the UK market, where the offset power considered forecasting errors in the system demand and generation. This strategy is claimed to enable better system planning to ensure that offset power benefits the battery owner without affecting the grid reliability.

Megel et al [114] explored the ability of batteries to provide frequency response during fast and zero-mean frequency deviations, while using an offset power to pass the slower and biased deviations to other resources, allowing extended service periods. This strategy claimed lower costs for the battery owner as well as reliable grid stability for the system operator.

Zhai et al [115] presented a dynamic SoC management strategy based on price signals from multiple markets for a battery participating in frequency response and

sport market price arbitrage in Australia. Profit maximisation for the battery owner is prioritised. Using a dynamic SoC target is shown to be better than a fixed SoC target for these types of services.

3.4.2.3 Scheduling strategy

Mariaud et al [116] assessed the energy arbitrage and frequency response based revenue streams for a battery with power/energy ratio of 0.66 in a UK commercial building site with solar power generation. The battery SoC is maintained in the range 33%–66% for frequency response such that high and low response could be provided. Before the evening system demand peak hours, the battery is charged to 100%, during which period frequency response is not provided. During the peak hours, so as to avoid the peak consumption charges, the battery provides backup to the commercial building load by discharging completely. After the peak hours, the battery is charged to 33% to continue the frequency response service.

Greenwood et al [117] considered a consumer energy storage asset participating in multiple DR programs such as frequency response, reserves and peak shaving in the UK. A high SoC paradigm that prefers power-to-grid services and a low SoC paradigm that prefers power-from-grid services were used. Demand reduction for peak shaving is given the highest priority since it's a consumer energy storage asset. For this purpose, a multiple linear regression based forecast model is built to estimate the consumer load in advance. Other services are added as additional layers on top of the peak shaving service. Based on the SoC paradigm and the service considered, SoC is adjusted either before the service or after the service.

3.4.2.4 EV strategies

SoC management in the context of EVs based grid balancing have been explored in the literature [118][119][120]. Compared to standalone and consumer energy storage, the EV batteries have smaller energy capacities. Hence their SoC management strategies include the added constraints of availability and aggregation.

3.4.2.5 The need for a data-driven SoC management strategy

The SoC management strategies developed for BESS based DR found in the literature are observed to be rule-based. Such approaches may not always be applicable for the reliable operation of a BESS in ongoing DR programs. For instance, the scheduling strategy encourages many BESS owners to recharge at the same time, burdening the grid stability and also making them pay for the additional electricity consumption. The real-time offset strategy interacts with other grid balancing services and creates issues related to planning [113]. With the exception of deadband strategy, the existing SoC management techniques are seldom implemented for reliable operation of BESS in ongoing DR programs. The presented research develops and demonstrates a novel data-driven SoC management strategy based on the building load estimation model and frameworks of ongoing DR programs.

3.5 Conclusions

The thesis applies data-driven building load estimation modelling for application in DR related activities such as capacity estimation, performance evaluation and reliable operation. The literature reviewed in this chapter explores the diverse areas associated with the thesis research.

Initially, different building load estimation approaches such as physics-based, hybrid and data-driven models were reviewed. The capability of these models and their application in building load estimation studies were explored. These models were then evaluated against the requirements of the DR applications addressed in this research. Based on the selection criteria such as reliability, adaptability, computational efficiency, cost-effectiveness, potential for large-scale deployment and automation, data-driven models were identified to be more suitable in this context.

Since the thesis focuses on supervised ML for data-driven building load estimation modelling, different supervised regression algorithms and model calibration methods used in similar studies were reviewed in detail. The algorithms such as linear regression, artificial neural networks, support vector machines, decision trees, ensembles and deep learning were evaluated in the light of their applicability for the problem at hand. Model calibration methods based on hyperparameter optimisation (HPO) were critically evaluated. Some of the methods were identified to be extremely computationally demanding and hence not recommended to be used in the research. Random-search HPO gives more control to the modeller in terms of choosing the model complexity according to the computational resource availability and hence proposed for model calibration as part of the data-driven modelling.

Literature related to data-driven DR operational tasks such as capacity estimation, performance evaluation and reliable operation were explored. Studies on data-driven capacity estimation and performance evaluation were found to be limited and those available did not address the need for large-scale deployment. The thesis identifies and fills this gap through development of data-driven models and compare their performance with conventional models used in ongoing DR programs in different electricity markets.

Studies on SoC management strategies for reliable operation of BESS in DR programs were reviewed. The state-of-the-art strategies were identified to be rule-based and the need for a data-driven SoC management strategy was proposed, such that it benefits the grid as well as the consumer.

In summary, the detailed literature review in this chapter has provided an overview of the related research that motivates the presented thesis. Different approaches were compared, contrasted and where relevant, research gaps have been identified.

Chapter 4

Data-driven building load estimation modelling

4.1 Introduction

In this chapter, a data-driven model is developed for large consumer building load estimation. The modelling is performed based on supervised machine learning (ML) with detailed investigation of aspects related to data pre-processing, model development and deployment. For demonstration and experimentation purposes, a representative office building dataset is used. The building load estimation model developed in this chapter is considered for application in demand response (DR) operations in Chapter 5 and Chapter 6.

4.1.1 Background

Building load estimation is an important activity in demand-side management based on which the electricity consumption behaviour of consumers could be analysed. This is particularly of interest in demand response (DR) where buildings provide grid balancing services, directly or through a DR aggregator. Many large consumer buildings are already involved in providing DR in different electricity markets and there is an increasing need for additional participation. Hence, it is important to consider modelling approaches that are reliable, adaptable, computationally efficient, cost effective and deployable for building load estimation in DR applications.

A detailed review of different building load estimation models was given in Chapter 3 Section 3.2. Physics-based models have been used widely in the literature and demand detailed information such as weather, geography, building design, thermo-physical variables, load characteristics, occupancy status and operating

schedules towards providing highly accurate building load estimates [7]. Since the physics-based models are complex, model development and computational times are generally higher. These models would need to be manually calibrated to changes in building conditions such as the addition of a new load or refurbishment of the building envelope, requiring additional information and knowledge of building physical systems. This limits the model's adaptability to new data, particularly in a deployment scenario where business continuity is of the highest priority. In addition to these, for DR aggregators with many building consumers in their portfolio, the lack of access to building specific information such as thermo-physical variables and occupancy status makes it further difficult to rely on physics-based models.

Data-driven models are interesting for building load estimation, particularly those based on supervised ML. These models are easily adaptable to new data and hence useful in deployment scenarios. For data-driven models, computational efficiency is a design consideration that can be prioritised in the development stage. Because of these advantages, data-driven models are enticing for load estimation of large consumer buildings participating in DR programs. A detailed evaluation of the different building load models in this context can be found in Chapter 3 Section 3.2.4.

There is rather no universally best ML algorithm and a model's predictive performance on a given dataset primarily depends not only on the ML algorithm, but also on the hyperparameter values, the predictor variables, the sizes of training-testing sets, the data quality, among other factors.

Artificial neural networks (ANN), support vector machines (SVM) [70,71] and decision trees [72–74] are the most widely used supervised ML algorithms in data-driven building load estimation. Other supervised ML algorithms of interest include Gaussian processes [75,76], extreme learning machines [78,79], nearest neighbours [77], deep learning [80] and ensembles [121,122] such as gradient boosted trees (GBT) [123,124]. These were reviewed in Chapter 3 Section 3.3.

Hyperparameter optimisation (HPO) implemented in data-driven building load estimation models include: heuristic methods such as manual-search [60,86–89], grid-search [83,90–93], random-search [125] and metaheuristic methods such as evolutionary/genetic algorithm [94,95] and swarm algorithms [39–43]. These HPO methods were reviewed in detail in Chapter 3 Section 3.3.3.

Predictor variables used in the supervised ML based building load estimation models include 1) weather variables such as ambient temperature, humidity, 2) building variables such as internal temperature, occupancy status, operational schedule, and 3) temporal variables such as time-of-day. The sizes of training-testing sets used and data cleaning measures taken, varies across the studies.

4.1.2 Contributions

Features of the data-driven building load estimation modelling presented in this chapter are listed below with the contributions highlighted.

Modern data science has offered a structured approach towards data-driven modelling. This starts with tasks such as data collection, data cleaning, feature engineering, feature selection and data scaling, that are part of the broader data pre-processing stage. Following these, training-testing data preparation, selection of ML algorithms, HPO and model selection are performed as part of model development stage. On top of these, there exist diagnostic tools such as learning curves, that are used for decision making in the modelling process. Such a structured approach towards data-driven modelling is missing in many of the building load estimation studies. Also, it is observed there are inconsistencies between empirical ML and applied ML in building load estimation. For instance, some of the previous building load estimation studies [14,25], have not given sufficient attention to HPO prior to comparing the performance of different ML algorithms on the same dataset. Data-driven building load estimation modelling presented in this chapter attempts to fill these gaps.

Computational efficiency without compromising the model's reliability is considered as an important design criterion in the selection of different methods within the data pre-processing and model development stages. For instance, apart from their general applicability on building load data, selection of ML algorithms is motivated by the simplicity of their architecture and computational resource requirement. For this reason, algorithms such as ANN, SVM and GBT have been prioritised over deep learning. Similarly, among the available HPO methods, random-search HPO has been implemented due its judicious use of available processing power to help identify good hyperparameter values from a large search space for each ML algorithm and hence improve its predictive performance. The potential to perform multiple rounds of HPO towards fine calibration of the model instead of a computationally demanding grid-search is also demonstrated in this study. Despite its advantages, random-search HPO has not been widely adopted in ML based building load modelling, possibly because computational efficiency has not always a primary design criterion.

In this study, aspects related to deployment of the data-driven model in an industrial operational environment such as integration with the application area, data flow and predictive performance tracking are discussed. Since model deployment is prioritised, it has also influenced the selection of methods in the data pre-processing and model development stages. For instance, data collection focuses only on the predictor variables that are realistically accessible for a DR aggregator. Building variables such as internal temperature and occupancy status are seldom accessible from the large consumer buildings and hence not used in the modelling. Feature selection tries to minimise the use of predictors with inherent errors, such as weather forecasts and their derivatives (referred to as class II predictors in this study). Instead it encourages the use of lag values of weather variables (referred to as class I predictors) which are measured and not forecasted values. Cross-validation is performed based on a forward sliding window of training-testing sets, simulating the actual training and forecasting process after deployment. The study also experiments the effect of possible data gaps between the training-testing sets on

the model performance. Development of such deployment-centric data-driven building load estimation models are limited in the literature.

Setting up a data-driven pipeline involving data pre-processing, model development and model deployment stages is important to ensure that the modelling methodology is replicable on a large number of buildings. Building load modelling studies focussed on replicability of the methodology for large scale applications are limited. The presented research fills this gap by developing a modelling approach in this chapter that is applied later in different DR operational tasks in Chapter 5 and Chapter 6.

4.1.3 Layout

The rest of the chapter is structured as follows. Section 4.2 addresses the first stage of modelling that is, data pre-processing and its sub-stages such as data collection, data cleaning, feature engineering, feature selection and data scaling. Section 4.3 explores the second stage of modelling that is, model development. This section includes discussions on the training and testing methodology, the theoretical foundations of supervised ML algorithms, the hyperparameter optimisation methodology and the model calibration/selection/evaluation methodology adopted in the study. Section 4.4 discusses the final stage, model deployment, which is relevant in the context of integrating the developed data-driven model to a real-world application.

4.2 Data pre-processing

The first stage of data-driven modelling adopted in this study is data pre-processing. This involves five sub-stages: data collection, data cleaning, feature engineering, feature selection and data scaling. The data collected for supervised ML consists of *predictor variables* (referred to as *predictors* henceforth) that may help predict a *response variable* based on the information they carry. Since the model is only as good as the data that flow into it, data cleaning and feature engineering are

important tasks in the pre-processing stage. The most informative predictors selected for model development are also referred to as *features* and the selection process is called feature selection. Data scaling and normalisation also contributes to the ML process. These are discussed in detail further.

4.2.1 Data collection

The study aims to build a data-driven model to estimate large consumer building loads. Hence, building load is considered as the response variable in the supervised ML approach adopted. Timestamped meter data is usually recorded for the whole building load. Although scarce, load specific sub-meter data may also be obtained. It is preferable to collect at least one year's worth data so that the seasonal influences (if any) could be understood. The interval of meter data recording varies based on buildings and electricity regulations of the land.

For the purpose of demonstration, annual whole building load meter data (referred to as *meter data* henceforth) from an office building located in the UK (with temperate maritime climate) is collected at half-hourly resolution. The rest of the analyses in this chapter uses this dataset as a representative example of other large consumer building datasets.

The potential predictors of building energy consumption can be categorised into weather, building and temporal variables. These are elaborated below.

4.2.1.1 Weather variables

Most large consumer buildings consume energy for heating, ventilation, air conditioning (HVAC), lighting, appliances and processes (related to the building function). The reliance on electricity for meeting these energy demands is ever increasing. HVAC, lighting and process loads contribute a major share to the total electricity consumption of these buildings. These loads are influenced by the local weather: 1) changes in ambient temperature and humidity affect the thermal comfort range maintained inside the building, forcing the HVAC loads to mitigate the effect by consuming more energy; 2) wind increases the heat transfer rates to

and from the building exterior surfaces and as a result affects the HVAC consumption; 3) solar radiation reaching the building envelope influences the solar gains (heat gain from solar radiation) and lighting levels inside the building. In general, *weather variables* such as temperature, humidity, wind speed and solar radiation influence building energy consumption. The magnitude to which these local weather variables influences the loads also depends on the building envelope and its thermal mass [126].

Weather variables can be collected from reliable sources such as meteorological stations. If weather data is not available from the nearby stations, data based on reliable weather models may also be used. It is worth mentioning that, microclimatic conditions and local landscape influence building energy consumption to a large extent. Hence it is important to ensure that the weather data collected are representative of the building location. This is particularly important for weather variables such as wind speed that is highly influenced by the orography, vegetation and nearby buildings. Temperature variations are usually consistent within a considerable distance from the building location [127].

In this study, ambient dry-bulb temperature is collected at hourly resolution from a meteorological station within a 2 miles radius of the office building. Derivation of useful features based on the temperature variable is discussed as part of feature engineering in Section 4.2.3.

4.2.1.2 **Building variables**

Apart from the direct energy use, people, appliances, lighting and building processes add to its internal heat gain, which is mitigated through additional cooling energy consumption. Indoor temperature settings, occupancy status, lighting levels, humidity levels and processes (depending on the building function) that influence the energy consumption from inside are considered as *building variables*. These are either not sufficiently monitored or not accessible due to security reasons in most large consumer buildings. With growing interest in smart buildings and sensors as well as increased data security measures, such data may become

more available in the future. In this study, no building variables are collected from the office building.

4.2.1.3 Temporal variables

Energy consumption patterns in many large consumer buildings are affected by *temporal variables*. For example, occupancy in an office or a supermarket varies across a day or a week based on the work routines and this results in load variations that are periodic. Diurnal changes (variations during the day) in ambient temperature and solar radiation also introduce periodicity to the building load variations. Many building processes are also dictated by time. Hence, timestamp recorded with the meter data is a potentially reliable predictor for building load. However, timestamp in the raw format cannot be used for ML modelling. Useful temporal features such as time-of-day will need to be derived and these are discussed as part of feature engineering in Section 4.2.3.

4.2.2 Data cleaning

Data collated from the real world are found to have outliers and missing values (gaps). Certain ML algorithms are inherently resilient to outliers and gaps. For example, the prediction equation used in the decision tree (DT) algorithm is a set of logic statements such as 'if predictor A is greater than X, the response is Y'. As a result, outliers do not usually influence the DT model performance. The DT algorithm also ignores missing data while building trees (more discussion in Section 4.3.2.3). In contrast, the predictive performance of algorithms such as MLR and ANN are highly influenced by the data quality [50]. In order to ensure a just comparison of predictive performance of different ML algorithms (as prioritised in this study), it is important to clean the data in the pre-processing stage. The major data cleaning activities are discussed below.

4.2.2.1 Outlier removal

Outliers are the samples that appear to deviate markedly from or are inconsistent with other samples in a dataset [128]. Figure 33 highlights the outliers in a sample of

the meter data collected from the office building. These may have resulted from metering errors.

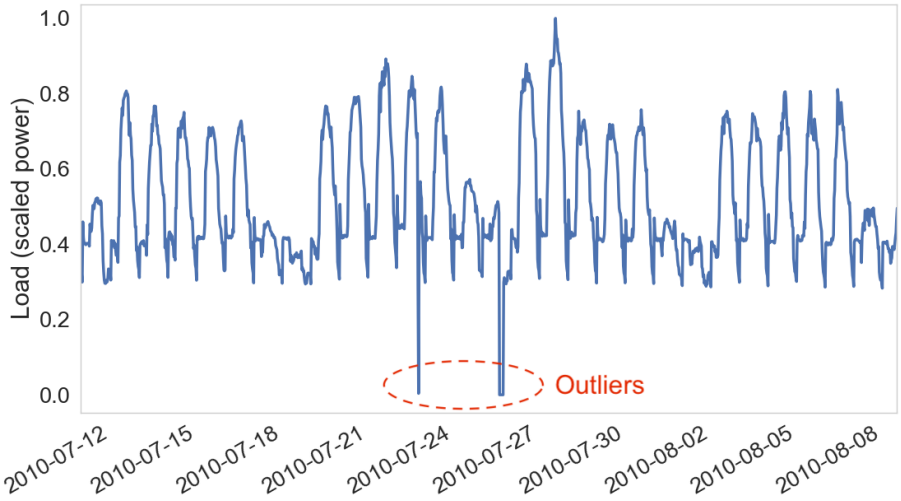

Figure 33: Office building meter data with outliers highlighted

Methods employed for outlier detection are categorised into parametric and non-parametric. **Parametric methods** are useful when the data distribution is already known or could be assumed, following which the parameters such as mean (μ) and standard deviation (σ) are derived. Samples that deviate from such a distribution are flagged as outliers. **Non-parametric methods** do not make any assumptions of the underlying data distribution [129][49].

Figure 34 shows the non-Gaussian nature of the office building meter data distribution. Many parametric methods for outlier detection assume the data to be Gaussian distributed [129] and hence would yield biased parameter (μ, σ) estimates for the meter data distribution. Further, in a data-driven context where new data samples are used to train an online deployed model, the underlying distribution of the data is susceptible to changes and new parameter estimates will need to be derived. For these reasons, parametric methods of outlier detection are considered to be less interesting for the model and brings our focus to the non-parametric methods.

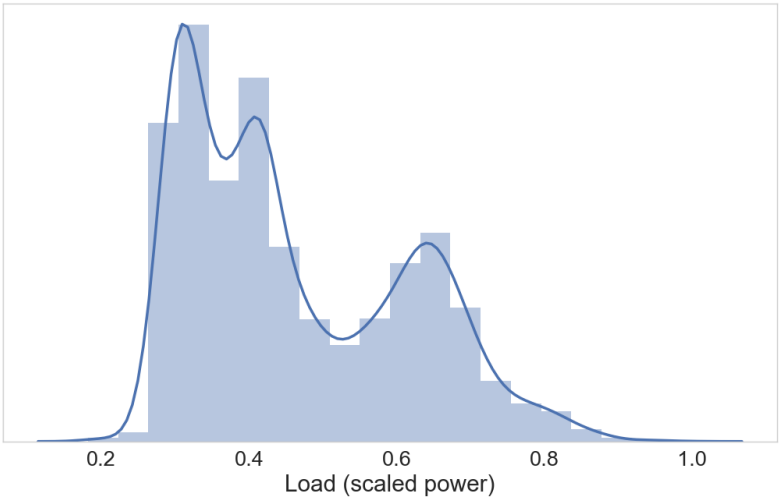

Figure 34: Non-Gaussian distribution of one year long half-hourly office building load whole meter data revealed by a histogram (with 20 bins) and a probability density function (blue curve)

A simple example of a non-parametric outlier detection method is the *Tukey fences* [130]. According to this method, for a dataset with Q_1 as the lower quartile (25th percentile), Q_3 as the upper quartile (75th percentile) and $(Q_3 - Q_1)$ as the interquartile range, the data samples outside the following range (for $k = 1.5$) are identified as the outliers: $[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$. Outliers need not always be extreme values, in which case Tukey fences method may not effective. In such contexts, ML based non-parametric methods such as two-class classification [128] and one-class classification [131] are promising. These methods also align well with the data-driven philosophy adopted in the presented research and are explored below.

Supervised two-class classification algorithms require training data with large number of outliers that are labelled into the following two classes: 1) a relevant class and 2) an irrelevant class (outliers). The training data should also cover the entire distribution to enable generalisation by the algorithm so as to avoid misclassification of unseen data points [128]. Based on the observation of meter data for large consumer buildings, we have concluded that the number of outliers is not typically large. Further, the distances of these outliers from the relevant data are not

consistent. Hence, it is not an easy task to identify and label them for supervised training. Multiple studies [132][133][134] have shown that, for problems with insufficiently labelled data or limited number of data samples of one class, the unsupervised one-class classification algorithm is found to be more effective than the supervised two-class classification.

With one-class classification there is no requirement for data labelling and a boundary for the relevant data is inferred from the training data samples. Any new/unseen data sample within this boundary is considered relevant and those outside are detected as outliers regardless of their distance from the boundary. The one-class classification is also observed to be more efficient than the two-class classification, since only the relevant data is required to infer the boundary [135]. For these reasons, a one-class classification algorithm is employed for outlier detection in the data-driven building load estimation model.

One-class classification is implemented using the one-class SVM algorithm introduced by Schölkopf et al. [131]. It is an adaptation of the SVM algorithm discussed in Section 4.3.2.2, for one-class classification problems. The parameter ν in one-class SVM corresponds to the probability of finding an outlier outside the boundary of the relevant data. The smaller its value, the lesser is the number of detected outliers. The tuning of parameter ν is done based on a manual-search of possible values because there are no specific rules that guide the selection of outliers, which is at the discretion of the modeller. For this purpose, the annual timeseries data is plotted and different values of ν are tried until the boundary of the relevant data discards what are considered to be outliers by the modeller.

Figure 35 shows the office building meter data for a period of 6 months. The outliers have been detected based on a one-class SVM with $\nu = 0.0085$. It is expected that the one-class SVM has learnt the boundary of relevant data and will continue doing the same after deployment.

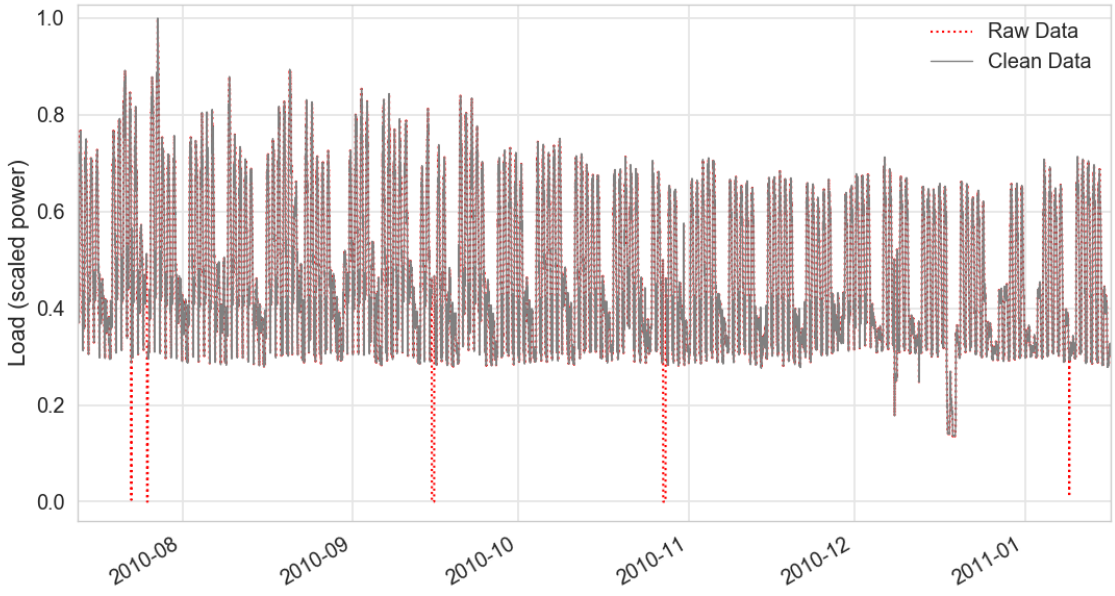

Figure 35: Outliers are detected from the raw meter data (red dotted line) using one-class SVM algorithm and are removed to produce the clean data (grey line)

4.2.2.2 Missing values

Gaps due to missing values do not occur randomly across the data samples used for model development. They are usually concentrated within specific variables. It is important to understand the reason behind the gaps in each variable on a case-by-case basis. Gaps may either be present during data collection or may occur as a result of outlier removal discussed previously. If the missing values are related to the response variable being predicted, such gaps are informative and should not be ignored [50]. If the percentage of missing values are substantially large (resulting in a sparse dataset), the variable may not be useful enough for model development. However, in many cases, gaps are small and the missing values can be replaced by imputation.

In the building load estimation model, single missing values of the predictor and response variables are imputed with the mean of the preceding and succeeding values in the timeseries. Data samples (rows) corresponding to variables (columns) with more than one consecutive missing values are removed prior to training the ML algorithm.

4.2.3 Feature engineering

Data in their raw form may not necessarily be informative to the ML model. Depending on the context, useful information needs to be extracted from the raw data. For example, instead of using timestamp as a predictor, an informative feature such as time-of-day can be extracted from the timestamp and used in the model. Further, new features may be derived from the raw data to improve the ML algorithms' predictive performance. For instance, polynomial transformation of a predictor into higher orders may help capture its non-linear relationship with the response variable. This is specifically useful for linear regression algorithm that cannot otherwise capture the non-linearity in the data [48]. In certain cases, meaningful domain-specific features may also be derived. In general, data transformations motivated by the type of data, the ML algorithms used and the domain knowledge of the modeller are referred to as feature engineering.

In the building load estimation model, feature engineering is performed on the temporal and weather variables to derive useful predictors. These are elaborated below.

4.2.3.1 Temporal predictors

Timestamps in the meter data are accessed in the format: *Year/Month/Day Hours:Minutes*. Temporal predictors such as time-of-day, day-of-week and month-of-year can be derived from these timestamps. For instance, the values for time-of-day ranges from integers 1 to 24, day-of-week ranges from integers 1 to 7 and so on.

In order to visualise the influence of temporal predictors on the building energy consumption, load profiling is performed on the office building. Figure 36 shows the influence of time-of-day, day-of-week and month-of-year on the building load. The curves represent annual mean values of loads over the time-of-day, across the day-of-week and month-of-year. The band around each curve represents the 95%

bootstrap confidence interval. A high level description of bootstrap confidence interval is given in the footnote[1]; readers may refer to [136] for further details.

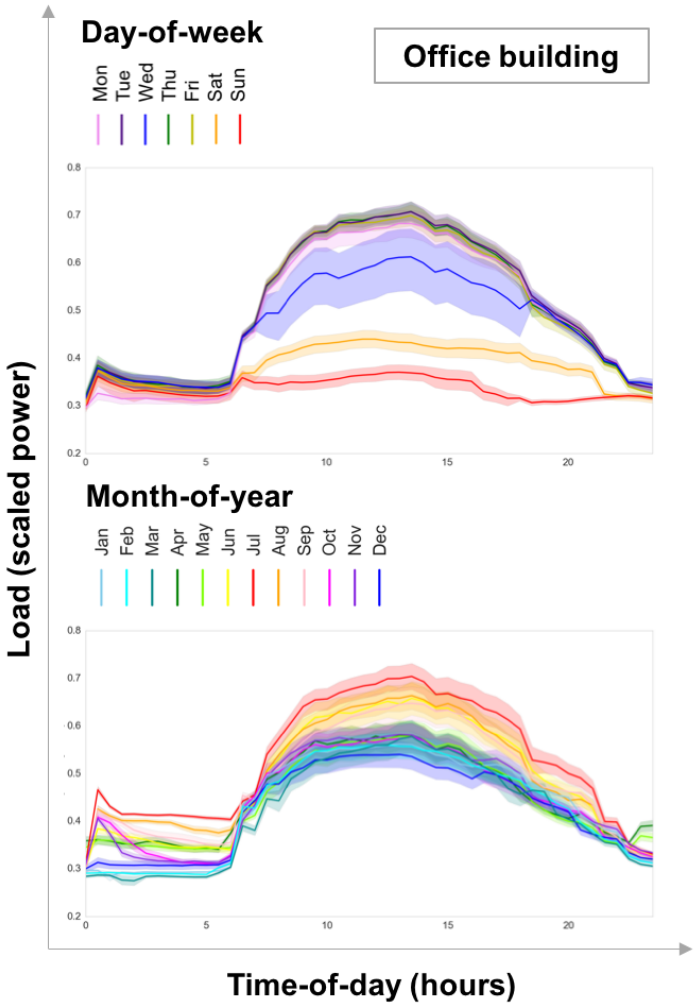

Figure 36: Variations of the office building load (bands show 95% bootstrap confidence interval)

[1] **Bootstrap confidence interval**: Data points are randomly sampled with replacement from a given distribution of size n to form an empirical bootstrap sample of the same size. The number of empirical bootstrap samples derived from the given distribution equals the number of preselected iterations i (in the above case $i = 10000$). The mean of the given distribution (x) is subtracted from the mean of each empirical bootstrap sample ($\mathbf{x}_i'$) to derive a distribution $\boldsymbol{\delta} = \mathbf{x}_i' - x$. Relevant percentiles of the $\boldsymbol{\delta}$ distribution are calculated for the required bootstrap confidence interval. In the case of the building load data, 95% bootstrap confidence interval is the range $[x - \delta_{0.025},\ x - \delta_{0.975}]$, where $\delta_{0.025}$ is the 2.5th percentile and $\delta_{0.975}$ is the 97.5th percentile.

Over the time-of-day, building load peaks at noon due to increase in cooling requirements. There is a consistent base load outside the working hours of the office. The time-of-day predictor will capture these diurnal load variations of the office building. Variations of the load across day-of-week are also significant. The building occupancy and subsequently the energy consumption during Sundays and Saturdays is lower compared to other days of the week. For some reason, the peak loads on Wednesdays are lower than the other working days. These operational aspects of the office building can be captured by the day-of-week predictor. Seasonal effects are evident from the load variations over the month-of-year. Higher loads are observed during the summer months due to increased cooling requirements. Meter data measurements for at least two years are required to capture this effect in a single variable such as the month-of-year or the week-of-year. Since such long-term data are not always available, other measures can be taken to capture the seasonal trends. These are discussed in the subsequent sections.

4.2.3.2 Weather predictors

As discussed in Section 4.2.1.1 weather variables such as temperature, humidity, wind speed and solar radiation influence the building energy consumption. Feature engineering performed on weather variables is based on the domain knowledge gathered from building energy research. Ambient dry-bulb temperature is the only weather variable collected for the office building load estimation. Derivation of useful predictors from the temperature variable are discussed below.

Degree days: Thermal comfort in a building is largely determined by the temperature and humidity levels maintained inside. In practice, HVAC systems operate when the internal temperature moves outside the thermal comfort range set by the thermostat [126]. Degree days are the summation of differences between the ambient temperature and a base temperature, where the latter is defined as the ambient temperature at which the HVAC systems do not need to operate in order to maintain comfort conditions. In the UK, a base temperature of 15.5 degree Celsius is widely used. When ambient temperature is below the base temperature, the heating

system provides heat proportional to the temperature difference. The heat energy consumption over a period of time relates to the summation of temperature differences and is referred to as the *heating degree days* (HDD). Similarly, cooling systems operate when the ambient temperature is above the base temperature, and the summation of their differences over a period of time gives the *cooling degree days* (CDD) [137]. HDD and CDD are good indicators of building heating and cooling energy consumption; hence these are proposed to be used as weather predictors in the building load estimation model. Depending on the availability of ambient temperature data and the use case, different methods are employed to calculate the degree days. Since the ambient dry-bulb temperature data are collected at hourly intervals, an hourly method is used for calculating the daily HDD (HDD_d) and daily CDD (CDD_d), based on the equations below:

$$HDD_d = \frac{\sum_{i=1}^{24}(T_b - T_i)^+}{24} \tag{4.1}$$

$$CDD_d = \frac{\sum_{i=1}^{24}(T_i - T_b)^+}{24} \tag{4.2}$$

where T_b is the base temperature, T_i is the ambient dry-bulb temperature at the i^{th} hour of the day and the plus symbol (+) highlights that the negative temperature differences are equated to zero [137]. Weekly and monthly degree days are calculated based on the summation of daily degree days over a week and a month respectively.

Estimates such as mean and extrema: It is possible to derive the mean, maximum or minimum (the last two referred to as extrema) of the weather variable over different time periods such as a day, a week or a month. These estimates are useful in capturing seasonal trends in the supervised ML model, particularly when long term data are not available. In the case of the office building, the daily and weekly mean/extrema of the dry-bulb ambient temperature can be used as predictors in the day-ahead and week-ahead[2] load estimation models, respectively.

[2] Forecast for a whole week in one go

Lag values instead of forecasts: Use of weather predictors for training ML algorithms brings in the responsibility of feeding forecasts in the online model once it is deployed. Any inaccuracy in these inputs will also be reflected in the estimated building load. The meteorological office in the UK claims that 92% of their day-ahead temperature forecasts are accurate within 2 degrees Celsius [138]. However, the impact of week-ahead weather forecast errors on the UK electricity demand estimates is significantly higher than that of the day-ahead forecast errors. In general, as the forecast horizon increases, the weather forecast accuracy declines [139,140]. In order to address this uncertainty, lag values of weather predictors such as the degree days and mean/extrema of the ambient temperature may be used over different time periods such as daily or weekly. For example, in the case of week-ahead building load estimation models, weekly mean/extrema temperature and weekly degree days from the previous week may be used during model development. Similarly, daily mean/extrema temperature and daily degree days from the previous day may be used for the day-ahead models. Such lag values can capture the seasonal trends without the uncertainty in using weather forecasts.

Further analysis on the selection of raw predictors and their derivatives for the day-ahead and week-ahead building load estimation models are performed in the next section.

4.2.4 Feature selection

Feature selection is the process of selecting a subset of the most informative predictors (referred to as features) based on relevance to the modelling problem. Feature selection is appealing because models with lesser number of features are computationally efficient and more interpretable. Non-informative predictors lower the predictive performance of parametrically structured (i.e. with fixed number of parameters that do not increase with training data) ML algorithms such as linear regression and artificial neural networks [50]. Feature selection can be performed using filter, wrapper or embedded methods. These are elaborated below.

Filter methods of feature selection are based on some criterion (such as correlation) that measures the relationship of predictors with the response variable. These methods are simple and independent of the ML algorithm. **Wrapper methods** identify the optimal combination of predictors that increases the overall model performance. The search for this predictor combination is part of the model training process making it computationally more demanding. **Embedded methods** of feature selection are specific to the architecture of a ML algorithm. For example, the decision trees algorithm has inherent feature selection capability and can produce feature importance values for the predictors used. Linear regression algorithms such as elastic net and ridge regression that penalise the regression coefficients of certain predictors also belong to this category. Embedded methods are computationally less demanding than the wrapper methods [141].

In this study, feature selection is performed based on filter methods. The strength of relationship between the predictor and the response variables are measured based on three different indices: *Pearson's correlation coefficient*, *Spearman's correlation coefficient* and *mutual information*. Pearson's coefficient measures the linear relationship between two continuous variables with the assumption of normal data distribution. On the other hand, Spearman's coefficient which is a non-parametric method is suited for non-normal data distribution, and additionally captures the non-linear relationship between two variables. The values of these correlation coefficients range in the interval [-1,1] where 1 represents the highest positive correlation, -1 represents the highest negative correlation and 0 represents no correlation. As a downside, Pearson's and Spearman's correlation coefficients can only measure monotonic relationships, where an increase or decrease in one variable results in an increase or decrease in the other [142]. Mutual information is a non-negative value which measures the amount of information one variable has about another. It can explain many complex relationships between two variables, including the non-monotonic relationship that the Pearson's and Spearman's correlation coefficients cannot [143]. The use of Pearson's correlation coefficient, Spearman's correlation coefficient and mutual information ensures that different

possible types of relationships between the variables are identified for selecting the features. All the correlation values (here correlation is used as a general term denoting strength of relationship between variables) are scaled to the interval [0,1] where 1 represents the highest correlation.

Feature selection using the above discussed indices is investigated on the office building dataset for a day-ahead and a week-ahead load estimation model. Predictors considered for the day-ahead model are listed in Table 1. Predictors with no uncertainty in the data are considered as Class I. These include temporal predictors such as time-of-day and lag values of ambient dry-bulb temperature and degree days. Those predictors with inherent errors (such as temperature forecasts and their derivatives) are considered as Class II. A similar classification is performed for the week-ahead model and the predictors are listed in Table 2.

Table 1: Predictors considered for the day-ahead office building load estimation model

	Predictors	**Description**
Class I predictors	time of day	Time of day with values from 0 to 24 in intervals of 0.5
	day of week	Day of week with values from 0 to 6
	daily HDD lag1	Daily heating degree days from previous day
	daily CDD lag1	Daily cooling degree days from previous day
	daily dtemp mean lag1	Mean of ambient dry-bulb temperature from previous day
	daily dtemp max lag1	Maximum of ambient dry-bulb temperature from previous day
	daily dtemp min lag1	Minimum of ambient dry-bulb temperature from previous day
Class II predictors	daily HDD	Daily heating degree days
	daily CDD	Daily cooling degree days
	daily dtemp mean	Mean of forecasted ambient dry-bulb temperature
	daily dtemp max	Maximum of forecasted ambient dry-bulb temperature
	daily dtemp min	Minimum of forecasted ambient dry-bulb temperature

Table 2: Predictors considered for the week-ahead office building load estimation model

	Predictors	Description
Class I predictors	time of day	Time of day with values from 0 to 24 in intervals of 0.5
	day of week	Day of week with values from 0 to 6
	weekly HDD lag1	Weekly heating degree days from previous week
	weekly CDD lag1	Weekly cooling degree days from previous week
	weekly dtemp mean lag1	Mean of ambient dry-bulb temperature from previous week
	weekly dtemp max lag1	Maximum of ambient dry-bulb temperature from previous week
	weekly dtemp min lag1	Minimum of ambient dry-bulb temperature from previous day
Class II predictors	weekly HDD	Weekly heating degree days
	weekly CDD	Weekly cooling degree days
	weekly dtemp mean	Mean of forecasted ambient dry-bulb temperature
	weekly dtemp max	Maximum of forecasted ambient dry-bulb temperature
	weekly dtemp min	Minimum of forecasted ambient dry-bulb temperature

Correlation of the predictors with the office building load is measured based on the indices such as Pearson's coefficient, Spearman's coefficient and mutual information. These are plotted in Figure 37 for the day-ahead model and in Figure 38 for the week-ahead model. The class I and class II predictors are grouped separately in both the plots.

Some of the general observations are: 1) the relationship between a predictor and the building load are measured differently by different indices; 2) correlation between the weather predictors and building load are stronger for the day-ahead models; and 3) the class I predictors (based on lag values of temperature) show similar correlation to the building load as the class II predictors (based on forecasts of temperature). Based on the 3rd observation, an experimentation is performed below, to assess whether there is any value in using class II predictors (which has inherent errors) for building load estimation.

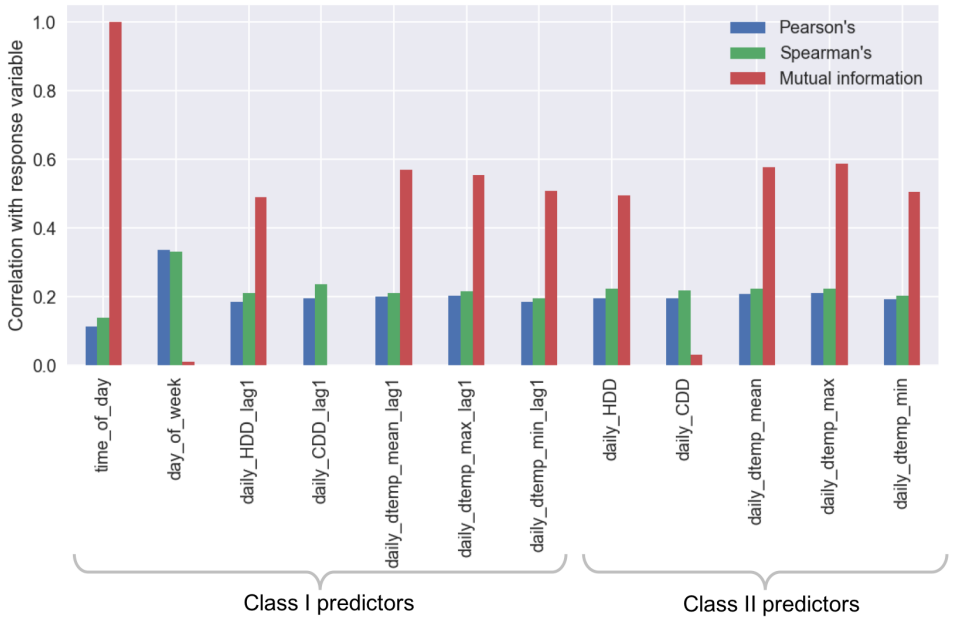

Figure 37: Strength of relationship between the predictors and building load for the day-ahead model

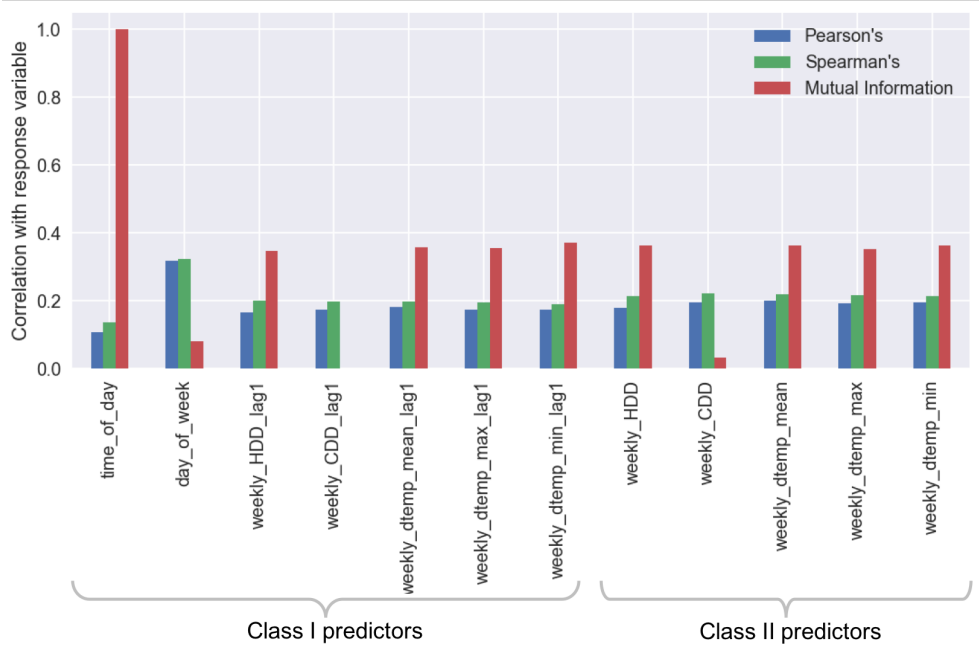

Figure 38: Strength of relationship between the predictors and building load for the week-ahead model

The experimentation takes insights from the model development discussed in Section 4.3, but is included here to keep the analysis in context. A decision trees algorithm (discussed in detail in Section 4.3.2.3) is used to build multiple supervised ML models by sequentially adding the predictors in decreasing strength of relationship as measured by mutual information. The one year long half-hourly data samples are split in the ratio 70:30 for training and testing. The initial model is developed using the most informative class I predictor, followed by the second model with the two most informative class I predictors and so on. Once all class I predictors are used up, class II predictors are added in the same order. Mean absolute errors (scaled) are measured for the building load estimates on the testing set for each of the model.

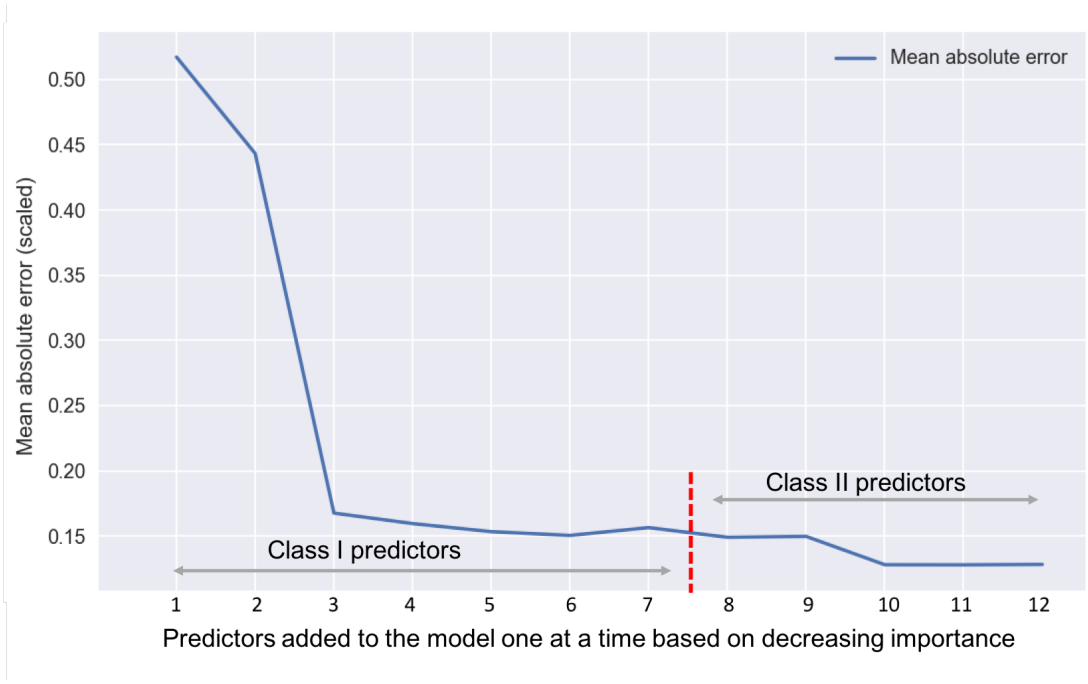

Figure 39: Class I and class I predictors added sequentially based on decreasing importance as measured by mutual information for day-ahead models

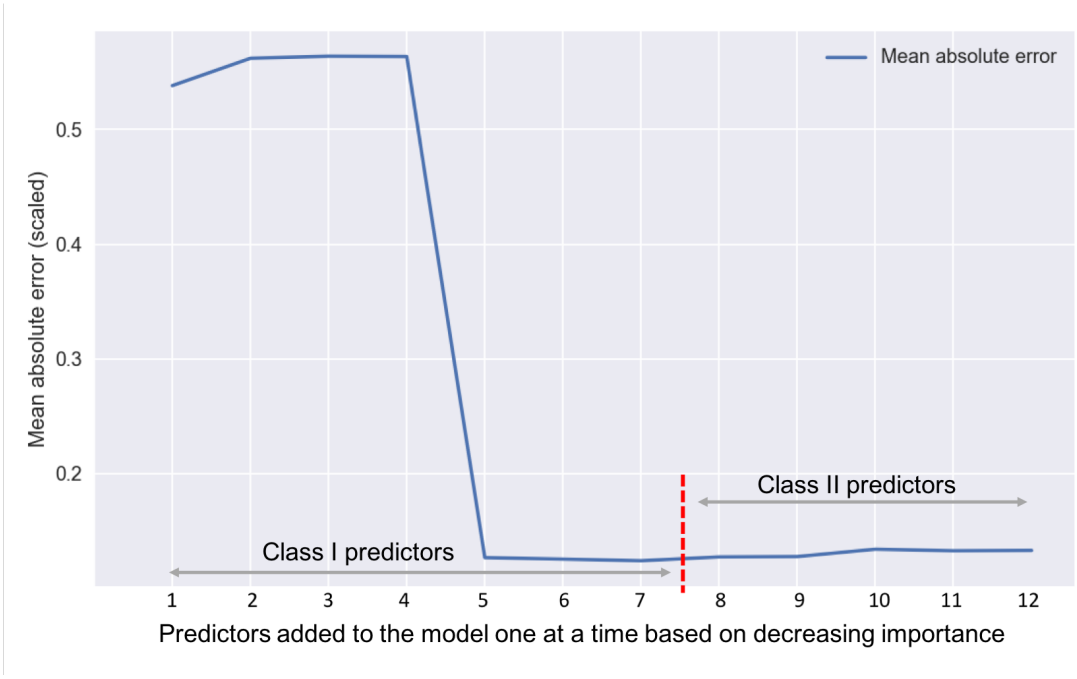

Figure 40: Class I and class I predictors added sequentially based on decreasing importance as measured by mutual information for week-ahead models

Figure 39 plots the errors for the day-ahead models and Figure 40 plots those for the week-ahead models. For the day-ahead models, inclusion of class II predictors does not provide any substantial decrease in the testing errors. Their inclusion in the week-ahead models are observed to increase the errors beyond what was achieved by the class I predictors. This may have occurred because the weekly estimates using forecasts (such as weekly HDD/CDD, weekly mean/extrema temperature) may not be representative of the weather influence on the week-ahead building load estimation. Based on this experimentation, the use of class II predictors is not recommended for the day-ahead and week-ahead building load estimation models. In addition to this, the use of class II predictors in a deployment scenario would mean that weather forecasts from third party services are collected with no prior knowledge of the forecast errors and this adds to the uncertainty in the building load estimates.

4.2.5 Data scaling and normalisation

Prior to model development, the data (predictors and response variable) can be transformed using scaling and normalisation to facilitate the ML algorithms' learning process. These are elaborated below.

Scaling transforms the variable data such that the values fit within a new scale. For example, *min-max scaling* transforms the range of raw data into the range [0,1], based on the formula given in Table 3. This method is robust to variable data with small standard deviations and preserves zero values if the dataset is sparse (i.e. having many zero values) [49]. Although scaling results in the loss of interpretability of the raw data, they help avoid the following issues resulting from features in different scales: 1) the learning process is dominated by the features with larger values, overlooking the information present in those with smaller values; 2) the gradient descent optimisation used in many supervised ML algorithms takes longer time to reach the optimal solution. In addition to this, scaling is also important for ML algorithms such as support vector machines and nearest neighbours that measure the distance to data points during training [50]. Hence, in the context of comparing the predictive performance of different ML algorithms, it is important to scale the variable data first.

Normalisation is performed to transform the variable data such that it could be described as a Gaussian distribution. This is important for training ML algorithms such as multiple linear regression and Gaussian processes that assume data to be normally distributed. In practise, the shape of the raw data distribution is often ignored, and a standardisation method is used to transform the data into zero mean and unit variance. The standardisation formula is given in Table 3. Here, the zero mean is achieved by subtracting the values with their mean, a task also referred to as centring. The unit variance is achieved by dividing the values with their standard deviation, which also helps in scaling the data. Standardisation is not very effective in the presence of outliers [49].

In the building load estimation model, since the outliers are removed prior to data transformation, each variable is scaled and normalised using the standardisation formula. The standardisation is applied consistently on the variable data during the development and deployment of the data-driven model.

Table 3: Scaling and normalisation methods

Methods	Formula*
Min-Max scaling	$\mathbf{x}' = \frac{\mathbf{x} - min(\mathbf{x})}{max(\mathbf{x}) - min(\mathbf{x})}$
Standardisation	$\mathbf{x}' = \frac{\mathbf{x} - mean(\mathbf{x})}{standard\ deviation\ (\mathbf{x})}$

*Here $\mathbf{x}$ is the raw variable dataset and $\mathbf{x}'$ is the transformed dataset

4.3 Model development

This section discusses the second stage of the data-driven modelling based on supervised ML, i.e. model development. Discussions on training and testing methods, architecture of ML algorithms, hyperparameter optimisation method, and model selection criteria have been included here.

4.3.1 Training and testing

The goal of a supervised ML model is to generalise beyond the data samples used to train the model and be able to predict on unseen test data. Training and testing on the same set of samples is a methodological mistake. Different aspects related to training and testing, such as predictive performance, bias-variance trade-off and cross-validation are discussed in this section. Experimentations are also performed towards data preparation for further model development.

4.3.1.1 Predictive performance

The supervised ML model learns a function from the data samples in the training set (that includes true values of the predictors and the response variable). Although

the model is trained using true values of the response variable, the function may not necessarily 'predict' them well. Comparing the predictions with the true values of the response variable in the training set gives the training error or training accuracy of the model. The same model is then used to predict the true values of the response variable in the unseen testing set, using true values of the features, based on which testing error or testing accuracy of the model is calculated. The predictive performance of regression models can be measured based on metrics such as coefficient of determination (R^2), mean squared error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), among others [144]. The selection of these metrics is problem dependant.

In this study, predictive performances of the supervised regression models are measured based on the R^2, MAE and MAPE metrics:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\mathrm{A}_i - \mathrm{E}_i)^2}{\sum_{i=1}^{n}(\mathrm{A}_i - \mathrm{A})^2} \tag{4.3}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\mathrm{A}_i - \mathrm{E}_i| \tag{4.4}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\mathrm{A}_i - \mathrm{E}_i}{\mathrm{A}_i}\right| \tag{4.5}$$

where E_i is the estimated value, A_i is the true value and A is the mean of all true values. Lower the MAE, better the model performance. The R^2 accuracy values in % are also used for measuring the model performance; in which case, higher the accuracy, better the performance.

4.3.1.2 **Bias-variance trade-off**

A supervised ML model with impressive predictive performance in the training set compared to that of the testing set may have learnt some unique but irrelevant characteristics of the training samples. Such a model is sensitive to small fluctuations in the data and is said to *over-fit* and have high *variance.* A model with low predictive performance in the training set may not have learnt the relevant

relationship between the features and response variable as intended. This model would give a low predictive performance in the testing set as well. Such a model is said to *under-fit* and have high *bias.* Higher the bias, lower the variance. The bias-variance trade-off is the problem of simultaneously minimising the prediction errors due to bias and variance so that the supervised ML model can better generalise on the unseen testing set [49]. The training and testing strategy adopted during model development determines the extent of these errors. This is where *cross-validation* plays an important role and is discussed in the following section. The concept of bias and variance are further used in inferring a *learning curve* discussed in Section 4.3.1.5.

4.3.1.3 **Cross-validation**

In practical applications such as demand response, the deployed building load estimation model needs to deliver reliable forecasts. Hence it is important to test the generalisation capability of the model to the best possible extent during model development. A validation set, in addition to the testing set, helps validate the predictive performance of the trained model. If the number of data samples used for model development are limited, it is not always possible to split them into training, testing and validation sets. In such cases, cross-validation (CV) is performed. An appropriate CV method also helps minimise the bias and variance errors of the model [50].

One of the widely implemented CV strategies for supervised regression is a **k-fold** method in which the original data samples are randomly split into k parts of roughly equal sizes. From these, k-1 parts are used for training a model whereas the held-out part is used for testing the model. This process is repeated with a unique held-out part each time, giving performance estimates for k different groups. For example, Figure 41 visualises a 3-fold CV where the data samples are split into three equal parts from which one part is always held-out for testing. **Leave-one-out** CV method is a variant of the k-fold method with k equal to the number of original data samples. For large datasets, this CV method is computationally demanding [49].

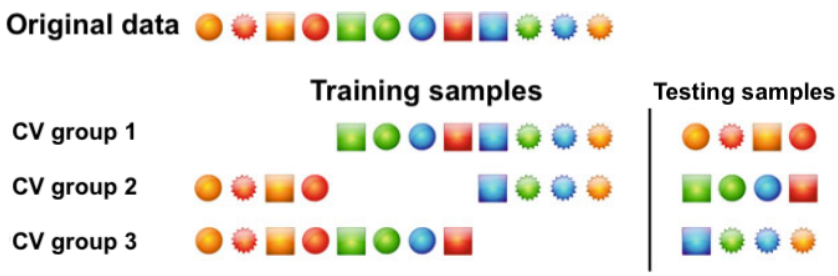

Figure 41: A 3-fold cross-validation [50]

The above-mentioned CV methods perform data shuffling. Depending on the application, the data samples may be split with or without shuffling. In this study, the office building dataset consisting of the samples of the features and the response variable (meter data) is partitioned to training and testing sets without shuffling, retaining the timeseries aspect as shown in Figure 42.

	Timestamp	Features (selected predictors)					Response variable
Training set	⋮					⋮	⋮
	t-5	X_1^{t-5}	X_2^{t-5}	X_3^{t-5}	....	X_p^{t-5}	Y^{t-5}
	t-4	X_1^{t-4}	X_2^{t-4}	X_3^{t-4}	....	X_p^{t-4}	Y^{t-4}
	t-3	X_1^{t-3}	X_2^{t-3}	X_3^{t-3}	....	X_p^{t-3}	Y^{t-3}
	t-2	X_1^{t-2}	X_2^{t-2}	X_3^{t-2}	....	X_p^{t-2}	Y^{t-2}
	t-1	X_1^{t-1}	X_2^{t-1}	X_3^{t-1}	....	X_p^{t-1}	Y^{t-1}
Testing set	t	X_1^{t}	X_2^{t}	X_3^{t}	....	X_p^{t}	Y^{t}
	t+1	X_1^{t+1}	X_2^{t+1}	X_3^{t+1}	....	X_p^{t+1}	Y^{t+1}
	t+2	X_1^{t+1}	X_2^{t+1}	X_3^{t+1}	....	X_p^{t+1}	Y^{t+2}
	⋮					⋮	⋮

Figure 42: Data samples partitioned to training-testing set retaining the timeseries aspect

At the time interval t, the value of the i^{th} feature is represented as X_i^t and the value of the corresponding response variable as Y^t. The number of features used in the load estimation model is given as p. The timeseries aspect helps take into account recent patterns in load variations during training. This may include recent changes to the building such as addition of a new load or refurbishment of the building envelope, with subsequent changes in energy consumption, to which we would want the model to adapt. Increase in the gap between training and testing sets may affect the predictive performance of the model. An experimentation is performed in Section 4.3.1.6 to explore and validate this assumption.

Since the timeseries aspect needs to be retained, partitioning of the training-testing sets for CV is performed on a **forward sliding window** basis. The training-testing sets window slides forward in steps equal to the size of the testing set. For demonstration, a sliding window with training set size of four weeks and testing set size of one week for the office building load is given in Figure 43. For this example, from a one year long dataset, the sliding window moving forward in one day steps will produce approximately 335 training-testing sets. These training-testing sets may then be shuffled prior to performing CV. The forward sliding window method of CV retains the actual training and forecasting patterns that would be conducted after model deployment.

Figure 43: Forward sliding window of training-testing sets

While the selection of an appropriate CV methodology is important, it is equally essential to determine the size of the testing set as well as the training set. These aspects are discussed further.

4.3.1.4 Testing set size determination based on application

Since the supervised ML model is aimed for deployment in DR programs, the size of the testing set is chosen based on the duration of the forecast period as required by the DR contract. Although these vary based on the DR program across the different electricity markets, we consider day-ahead and week-ahead forecast periods to be the most representative. Hence this chapter focuses on day-ahead and week-ahead building load estimation models only.

4.3.1.5 Training set size determination using learning curves

The size of the training set depends on the number of data samples required to train a supervised ML model such that it produces reliable building load estimates. *Learning curve* is a diagnosis method used in supervised ML to assess the impact of increasing training data samples on a model's predictive performance. The learning curve helps understand the errors due to bias and variance in the ML model through cross-validation (CV) [49].

Learning curves are generated using the office building dataset and 3 ML algorithms: Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Gradient Boosted Trees (GBT). These ML algorithms are discussed in detail in Section 4.3.2. Features such as time-of-day, day-of-week, daily degree days from previous day and daily mean/extrema ambient dry-bulb temperatures from previous day are used to train the day-ahead building load estimation models. Similarly, features such as time-of-day, day-of-week, weekly degree days from previous week and weekly mean/extrema ambient dry-bulb temperatures from previous week are used to train the week-ahead building load estimation models. These features belong to the class I category discussed in Section 4.2.4.

From the annual half-hourly office building dataset, training and testing sets are derived using the forward sliding window CV method (discussed in Section 4.3.1.3) such that the window slides forward: 1) in daily steps for a day-ahead model (with test set size of one day) and 2) weekly steps for a week-ahead model (with test set size of one week). For each of these models, training sets of sizes 1 to 24 weeks are considered with increments of one week and CV is performed using the 3 ML algorithms on 50 randomly selected training-testing sets. The averages of training and testing predictive performances obtained from CV are used to plot the learning curves. Mean absolute error (scaled) metric is employed for measuring the predictive performance. Figure 44 shows learning curves for the day-ahead models and Figure 45 shows those for the week-ahead models.

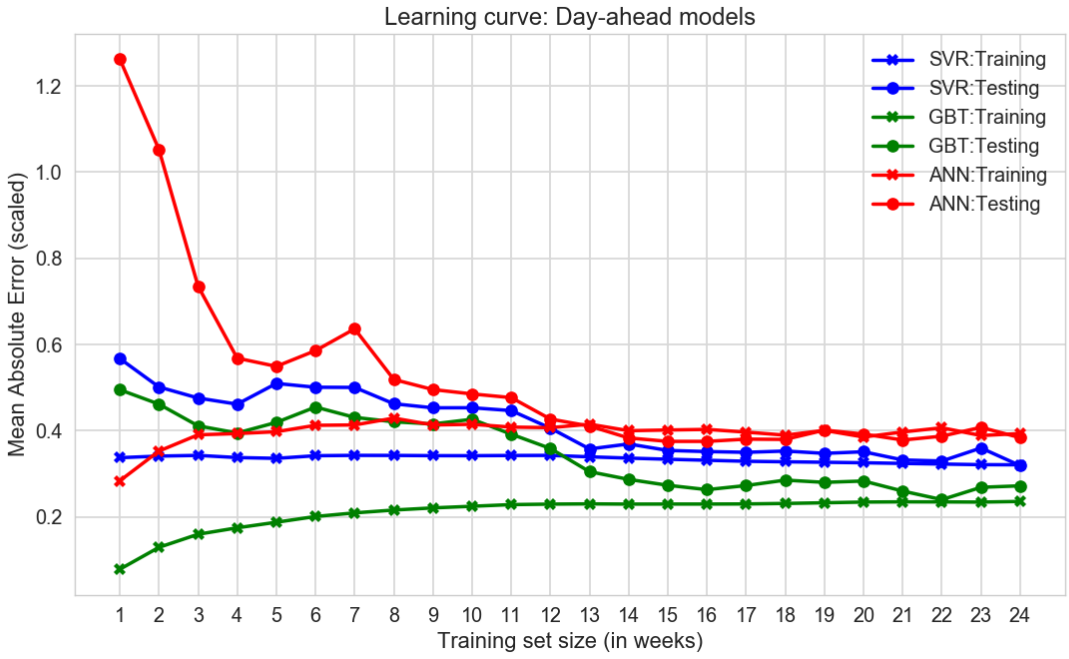

Figure 44: Learning curves for day-ahead office building load estimation models

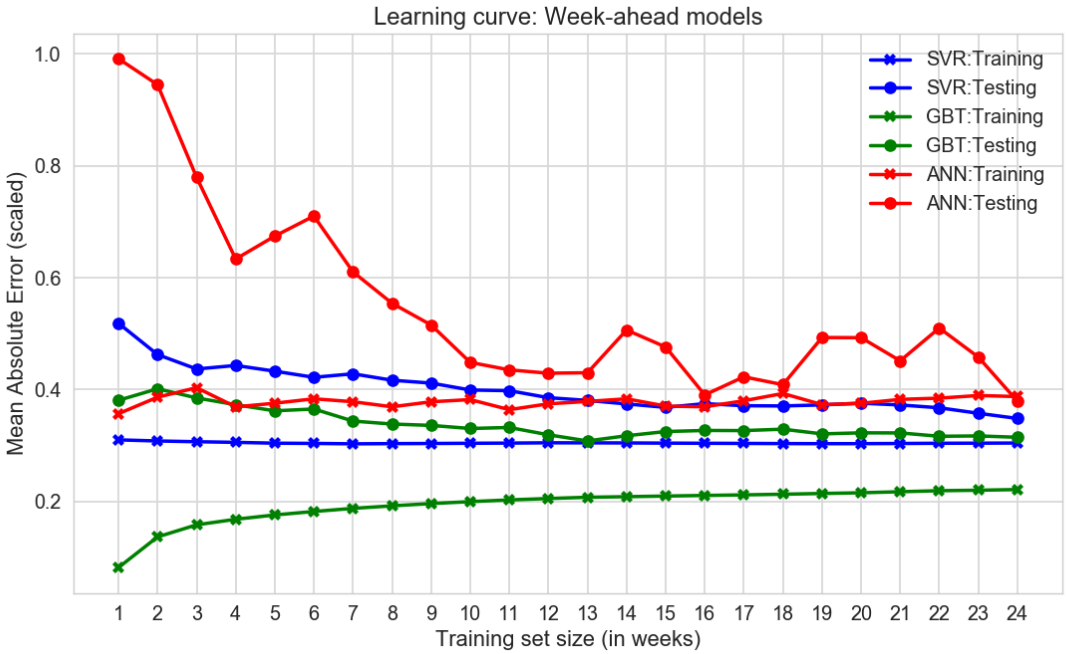

Figure 45: Learning curves for week-ahead office building load estimation models

Let us observe the general trends in the learning curves. For both the day-ahead and week-ahead models, when the training set size is the smallest (1 week), the average training error is the lowest and the average testing error is the highest. This is expected because the models overfit on the small training set and are unable to generalise on the testing set, denoting high variance. With increasing training set sizes, it is observed that the training errors increase and the testing errors decrease, showing signs of convergence. This shows that the model variance is being reduced, as a result of which its generalisation capability improves.

If the training and testing errors converge to a common point, there is no requirement to further increase the training set size. If they do not converge, adding more training data samples without any substantial decrease in testing errors would increase the computational cost of the model. Hence, it is recommended to stop increasing the training set size at this point. For the day-ahead and the week-ahead models, a comparison across different ML algorithms shows that the decrease in

testing errors are not substantial beyond the training set size of 13 weeks. Hence, this is fixed as the training set size.

ML algorithms used in this analysis have not been calibrated (using hyperparameter optimisation, described in Section 4.3.3) and hence do not showcase their best performances. However, a general comparison shows that bias errors due to GBT models are the lowest and those due to ANN are the highest. Predictive performance of the models could be further improved through calibration and this is explored in Section 4.3.4.

4.3.1.6 Effect of increasing gap between the training and testing sets

The strategy adopted so far makes use of training data samples directly preceding a testing set. In the deployment scenario, it may so happen that the data flow is interrupted and there are no sufficient preceding data samples for training. As a result, the day-ahead or week-ahead load estimations will have to be performed using models trained on the data samples from distant past.

Considering this issue, an analysis is performed to assess the effect of the gap between the training set and the testing set. For the day-ahead and week-ahead models, the training set sizes are fixed at 13 weeks (based on learning curves in the previous section). The gaps starting from 0 (no gap) up to 21 weeks are incremented in weekly steps. In each step, training-testing sets are prepared using the forward sliding window method, from which 10 sets are randomly selected for cross-validation. Average mean absolute error (scaled) on the testing sets are measured for the same 3 ML algorithms used in the previous section and plotted for the day-ahead (Figure 46) as well as the week-ahead (Figure 47) models.

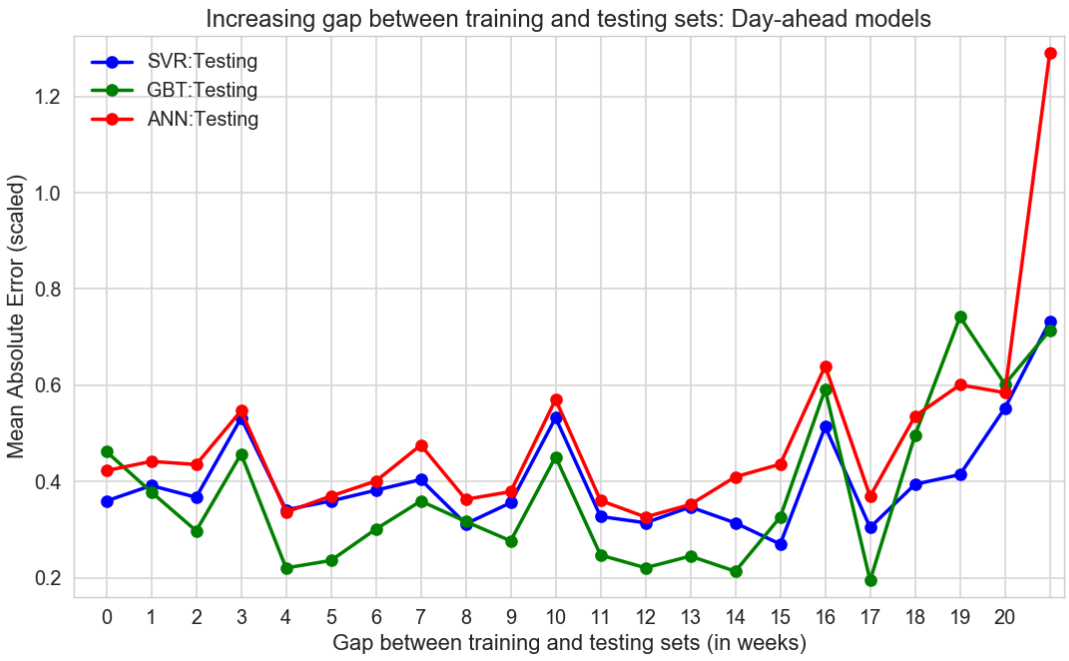

Figure 46: Effect of increasing gap between training and testing sets for day-ahead load estimation models

For gaps less than 18 weeks (~ 4.5 months), the average testing errors in the day-ahead models are observed to vary inconsistently within a range (0.2 to 0.6). If data gaps are expected in deployment, it is recommended to train different day-ahead models for different sizes of gaps. Compared to SVM and GBT, the errors in ANN based day-ahead models are higher. For gaps beyond 18 weeks, the testing errors increase significantly, making all the day-ahead models unreliable. In the supervised ML model developed, training sets are selected such that they directly precede the testing sets without any gaps, for all day-ahead models.

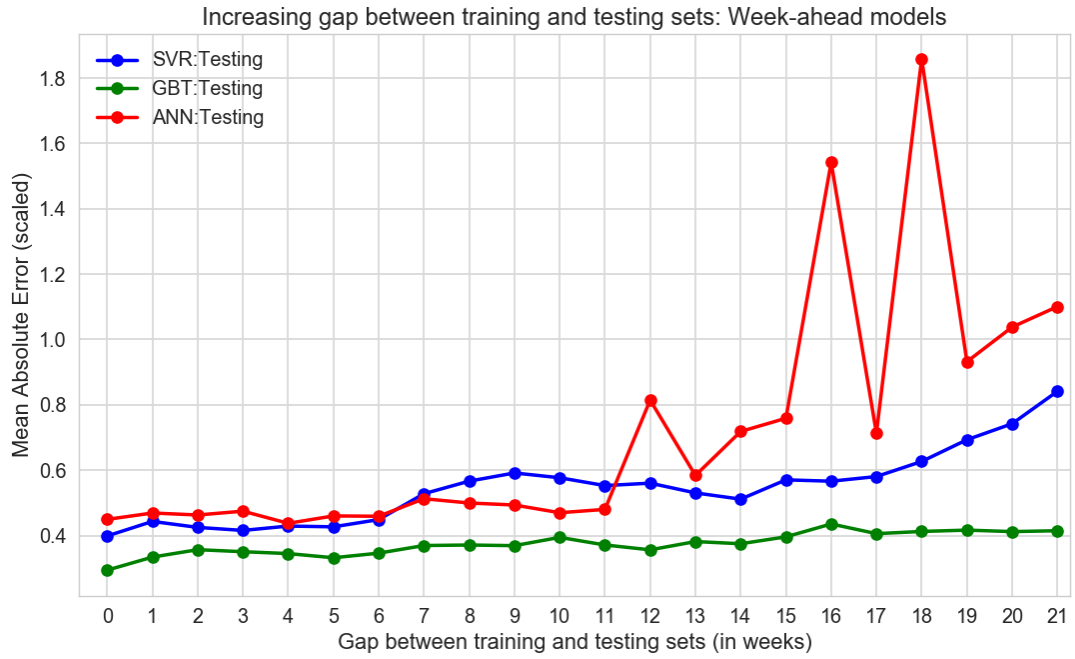

Figure 47: Effect of increasing gap between training and testing sets for week-ahead load estimation models

For week-ahead models, the testing errors are consistent and hence reliable up to a gap of 6 weeks. Beyond this timeframe, there is an increasing trend in the testing errors. Depending on the accuracy requirement of the application, marginal increase in errors may be tolerated. GBT based week-ahead models are observed to be the least sensitive to gaps, followed by SVM. The ANN based week-ahead models are the most sensitive to gaps, as noticed earlier for the day-ahead models. In the supervised ML model developed, training sets are selected such that they directly precede the testing sets without any gaps, for all week-ahead models.

4.3.1.7 Development and evaluation sets

Training and testing using CV helps select the supervised ML model with the best predictive performance. However, in order to ensure that the model is reliable in a deployment environment, it is useful to further validate its performance on other unseen data. For this purpose, the training-testing sets selected using a forward sliding window, without any gaps between them, are randomly sampled and

equally divided into a development set and an evaluation set. This can be visualised from Figure 48.

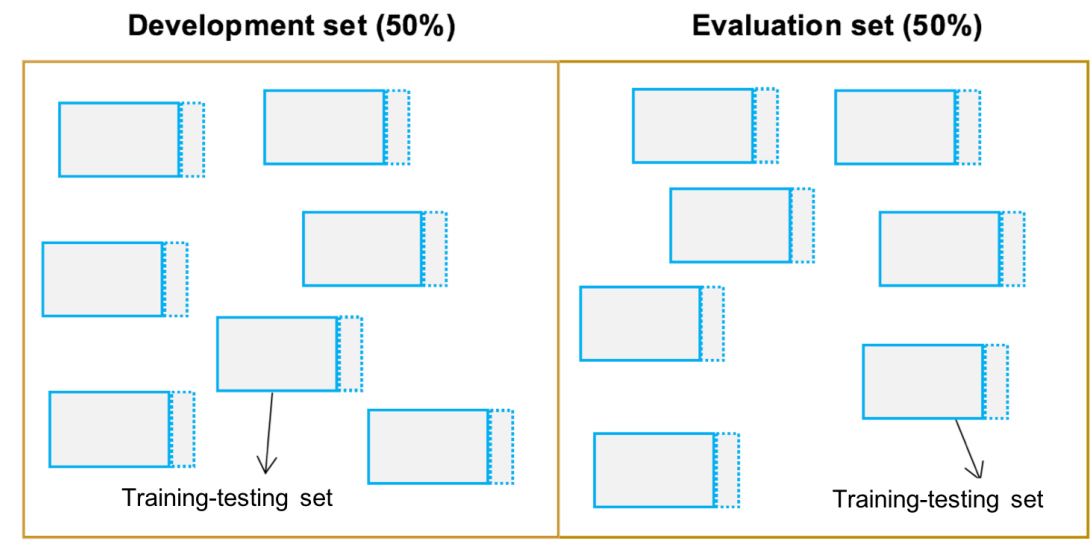

Figure 48: Training-testing sets randomly sampled into development and evaluation sets

The day-ahead and week-ahead models based on multiple supervised ML algorithms are calibrated using hyperparameter optimisation (HPO) in the development set and their best versions are further cross-validated in the evaluation set. The model with the best average testing error in the evaluation set is selected for final deployment. These aspects are discussed in detail as part of the model calibration, selection and evaluation in Section 4.3.4. Before that, ML algorithms, their hyperparameters and the HPO method used in this study are discussed in the following two sections.

4.3.2 ML algorithms and their hyperparameters

Among the different supervised ML algorithms developed by the empirical ML community, many were applied in building load estimation studies. These were reviewed in Chapter 3 Section 3.3. Although simple in architecture, linear regression algorithms cannot comprehend non-linear relationships between variables and building load without explicitly using feature transformations such as polynomials.

This issue was discussed in Chapter 3 Section 3.2.2.1.1. Since load characteristics are different for each building, it is not easy to identify the appropriate feature transformation (need not always be polynomial). Replication of this task on large number of buildings as required for DR applications becomes a tedious process. Based on this reasoning, linear regression algorithms are not adopted in this study. In contrast to the simple linear regression algorithm, deep learning algorithms are extremely complex. A comparison of the predictive performance of different ML algorithms in building load estimation has shown that the deep learning model achieves similar performance on a one year dataset, but using higher computational resources and complex training schemes [80]. Algorithms such as artificial neural networks (ANN), support vector machines (SVM), decision trees (DT) and ensembles are good with non-linear data [50], have proven predictive performances on building load data and are computationally less demanding than deep learning algorithms. Based on these advantages, ANN, SVM and a DT based ensemble algorithm namely gradient boosted trees (GBT) are chosen for further investigation in this chapter. The ANN and SVM are representative of single algorithms whereas the GBT is representative of an ensemble algorithm. Due to the computational complexity involved, the scope for investigating other ML regression algorithms such as Gaussian Processes [55] and nearest neighbours [56] has been limited in this study. Nevertheless, the developed methodology and experimentations are replicable on other supervised ML algorithms as well.

Architectures of the ANN, SVM and GBT algorithms along with their hyperparameters are elaborately discussed below. Their hyperparameters are given in *italics* for easy identification in the succeeding sections.

4.3.2.1 **Artificial neural network (ANN)**

The ANN is an ML algorithm inspired by the functioning of the human brain. The arrangement of neurons (or units) into layers and their interconnection determine the architecture of an ANN. Single-layer networks have an input layer (that receives the predictor variables) and an output layer (that delivers the response variable),

but no hidden layer between them. The weighted interconnect links between the input and output layer is considered as one layer and hence the name single-layer. Single-layer networks seldom have the capability to learn complex relationships between the input and output. For this purpose, multi-layer networks with one or more hidden layers between the input and output layers are used. They are particularly good with non-linear data [145].

Based on the information flow, neural networks are classified into: 1) feedforward networks – with no feedback connections and information flowing exclusively in the forward direction; and 2) feedback networks – having feedback connections and information flowing forward and backward. Figure 49 represents a multi-layer perceptron, which is a class of feedforward multi-layer neural network used for supervised regression.

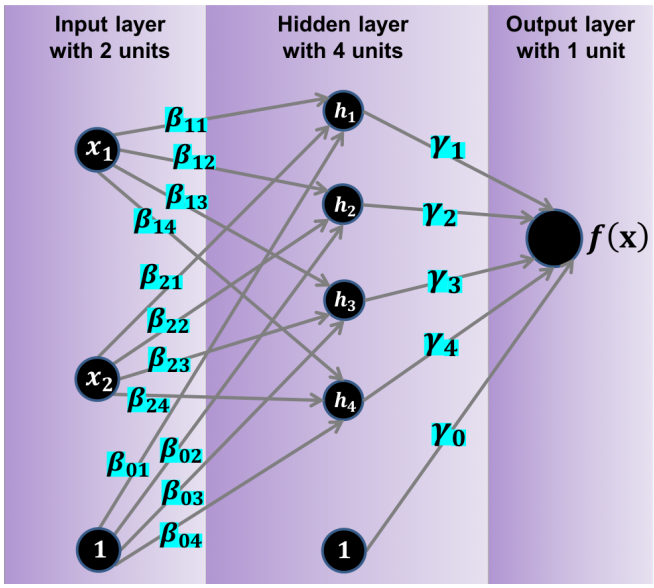

Figure 49: Representation of a multi-layer neural network with weights highlighted (in blue) and bias as a unit of value 1

The following mathematical formulation from [50] is given for a multilayer perceptron with one hidden layer. The input layer takes in P predictor variables. Each unit in the hidden layer is a linear combination of one or more of the predictor variables. This linear combination is transformed by an activation function such as

the step, logistic sigmoid, hyperbolic tangent (tanh), or rectified linear unit (relu). The same activation function is applied on units of the same layer (although this is not a mandate). Some of the activation functions used for supervised regression are listed in Table 4.

Table 4: Activation functions used in supervised ANN regression

Activation function	Formula	Remark
Logistic sigmoid	$g(x) = \frac{1}{1+e^{-x}}$	Outputs range from 0 to 1
Hyperbolic tangent (tanh)	$g(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$	Outputs range from -1 to 1
Rectified linear unit (relu)	$g(x) = \max(x, 0)$	Output is always positive

The k^{th} hidden unit is given as:

$$h_k(\mathbf{x}) = g\left(\beta_{0k} + \sum_{j=1}^{P} x_j . \beta_{jk}\right) \tag{4.6}$$

where, x_j is the j^{th} predictor, β_{jk} is the regression coefficient (or weight) of x_j on the k^{th} hidden unit and β_{0k} is the bias (weight of a unit whose value is 1) on the k^{th} hidden unit. In the case of neural networks with two hidden layers, units of the second hidden layer are estimated based on the weights from the first hidden layer, the activation function and the bias. With more hidden layers, this process repeats. However, we continue the discussion based on one hidden layer.

Since the weights are being summed, they should be on the same scale. Hence it is recommended that the predictor variable values are scaled prior to modelling. It is also preferable to perform feature selection prior to training since non-informative predictors increase the model complexity and lower the predictive performance. The response in the output layer is a linear combination of H hidden units in the previous layer as follows:

$$f(\mathbf{x}) = \gamma_0 + \sum_{k=1}^{H} \gamma_k . h_k \tag{4.7}$$

Here, γ_k is the weight on h_k and γ_0 is the bias. Figure 49 presented earlier highlights the weights between the input, hidden and output layers of a multi-layer perceptron.

For a neural network with one hidden layer, P predictor variables and H hidden units, the number of regression parameters (weights and bias) to be optimised are $H(P+1)+H+1$. Optimisation techniques such as the stochastic gradient descent (SGD), Levenberg–Marquardt (LM) algorithm or Limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm may be used for this purpose. Using the training-set of data with n samples, the regression parameters are optimised by minimising the loss function (least squares):

$$k(x) = \sum_{i=1}^{n}(y_i - f(x_i))^2 \tag{4.8}$$

where, y_i is actual value of the response variable and the $f(x_i)$ is the estimated value. Since the number of regression parameters to be estimated are large, the neural networks have the tendency to over-fit the predictor-response relationship. One of the techniques to tackle this is, by using the following cost function that penalises the regression parameters:

$$k(x) = \sum_{i=1}^{n}(y_i - f(x_i))^2 + \alpha.\sum_{k=1}^{H}\sum_{j=1}^{P}{\beta_{jk}}^2 + \alpha.\sum_{k=1}^{H}{\gamma_k}^2 \tag{4.9}$$

where, α is called the regularisation term. Higher the value of α, smoother the fit on the training-set.

The multilayer perceptron is trained using a backpropagation method which involves three phases: 1) the feedforward of the input training pattern, 2) the backpropagation of the cost function (or error) and 3) update of the weights to minimise the cost function [145]. Figure 50 is a graphical representation of the backpropagation method on a multi-layer perceptron with one hidden layer. After the neural network is trained, only the feedforward phase is implemented for the testing on unseen data. Hence testing is always faster than training.

Readers may refer to [46,59,146,147] for more discussions on ANN for supervised regression.

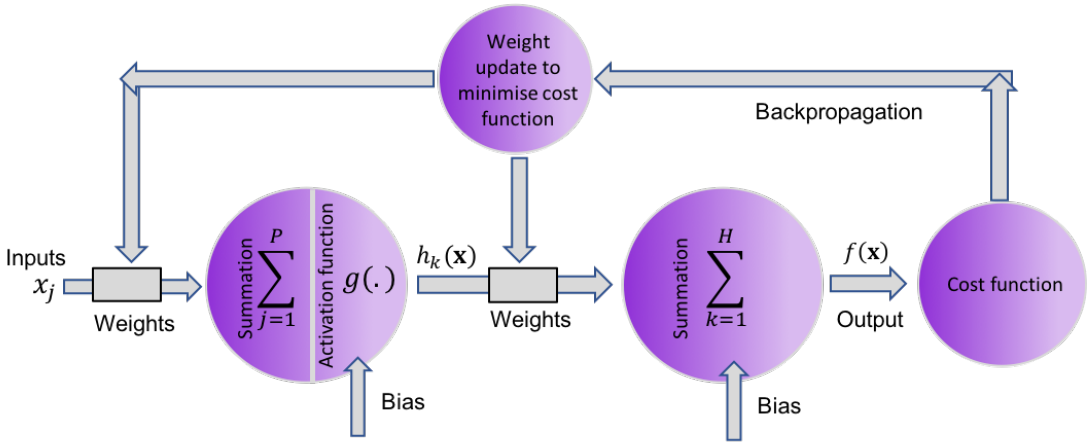

Figure 50: Graphical representation of backpropagation in a multi-layer perceptron with one hidden layer

The hyperparameters of ANN algorithm considered in this research are the *hidden layer size* (number of hidden layers and numbers of hidden layer neurons), the *activation function*, the *regularisation term* and the *optimisation technique*.

4.3.2.2 Support vector machines (SVM)

SVM's were originally conceptualised in the context of classification and later expanded to regression problems. The mathematical formulation for SVM regression in this section is adapted from [50].

For an unseen data sample **u** with P predictor variables, a multiple linear regression (MLR) model predicts the response variable based on the regression parameters **β** as follows:

$$f(\mathbf{u}) = \beta_0 + \beta_1 u_1 + \beta_2 u_2 + \cdots \ldots \ldots + \beta_P u_P = \beta_0 + \sum_{j=1}^{P} \beta_j u_j \qquad (4.10)$$

An SVM regression model predicts the response variable using regression parameters estimated as a function of a set of unknown parameters α_i and n training set data samples x_{ij} as:

$$f(\mathbf{u}) = \beta_0 + \sum_{j=1}^{P} \sum_{i=1}^{n} \alpha_i x_{ij} u_j \tag{4.11}$$

The SVM regression differs from the MLR regression in the following aspects: 1) SVM has as many α parameters as there are training-set samples and 2) SVM uses a subset of the training-set data x_{ij} in the predictions.

For the training-set samples within $\pm\epsilon$ (epsilon) of the regression line (called the insensitive tube), the α values are set to zero and they do not influence the prediction. As a consequence, the subset of the training-set samples with $\alpha \neq 0$ determine the regression line and they are called the support vectors. Figure 51 is a residual versus predicted value plot showing the support vectors (grey circles) for an SVM with $\epsilon = \pm 0.01$ (blue lines).

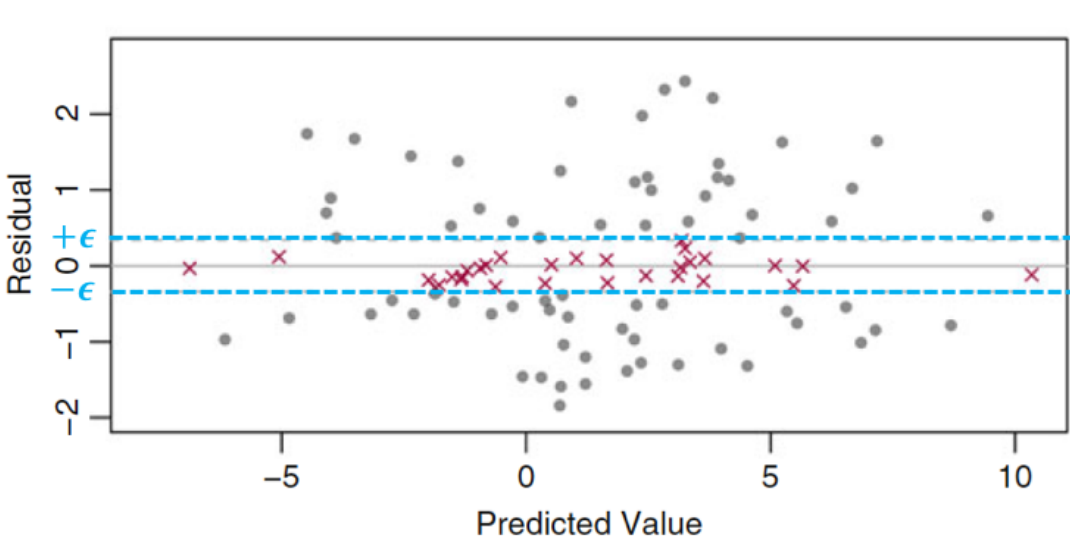

Figure 51: Support vectors (grey circles) and other samples (red circles) of an SVM regression model with $\epsilon = \pm 0.01$ (blue lines) shown using a residual vs predicted value plot (adapted from [50])

Based on two simulated datasets, Figure 52 shows that the linear regression lines are influenced by outliers, while SVM proves to be more robust. This is because, SVM avoids the use of squared errors that magnify the presence of outliers.

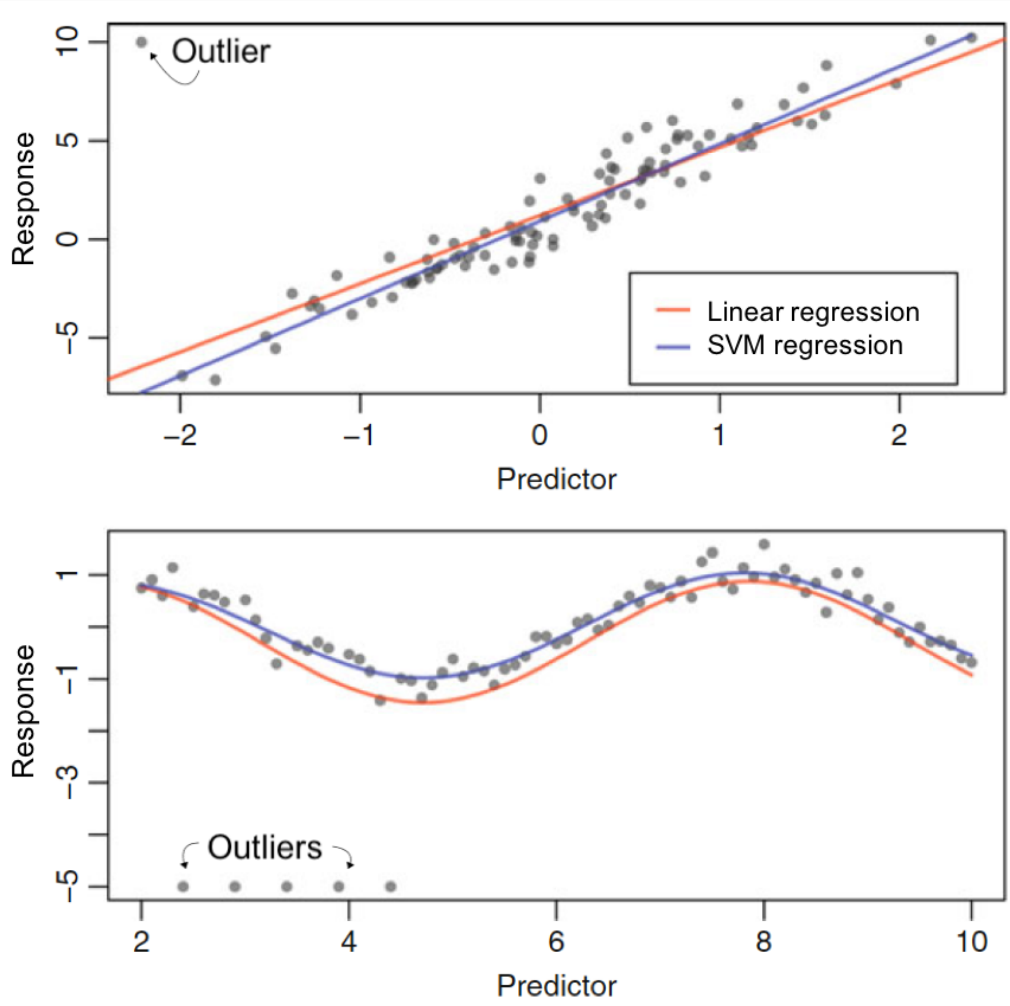

Figure 52: Outlier robustness of SVM in comparison to linear regression, based on a linear dataset with one large outlier (top) and a sin wave dataset with several outliers (bottom) (adapted from [50]).

The equation (4.11) could be rewritten as:

$$f(\mathbf{u}) = \beta_0 + \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{P} x_{ij}\, u_j$$

$$= \beta + \sum_{i=1}^{n} \alpha_i K(\mathbf{x}_i, \mathbf{u}) \qquad (4.12)$$

where $K(\mathbf{x}_i, \mathbf{u})$ is the dot product $\mathbf{x}_i'\mathbf{u}$. Here $K(.)$ is a linear kernel function that maps the predictors into a high-dimensional feature space with the help of support vectors. This is generalised to non-linear kernel functions (referred to as the 'kernel trick') such as polynomial, hyperbolic tangent (sigmoid) and radial basis function (RBF). The right choice of a kernel function and its corresponding parameters will enable the separation of the data by a hyperplane. The non-linear SVM regression fit in the sine wave dataset of Figure 52 is based on a model with RBF kernel.

Table 5 lists the kernel function formulae. The γ value plays an important role in the performance of the kernel and needs to be selected carefully. If overestimated, the kernel will start to lose its non-linear power and if underestimated, the model will be highly sensitive to noise in the training data (high variance) [148]. The kernel function parameters such as *degree* and γ need to be pre-set by the modeller, and hence are considered as hyperparameters.

Since differences in predictor scales could affect the model, it is recommended to standardise the data prior to modelling. Feature selection prior to training is also recommended since the non-informative predictors may lower the predictive performance of this parametrically structured algorithm.

Table 5: Kernel functions used in supervised SVM regression

Kernel function	Formula
Linear	$K(\mathbf{x}_i, \mathbf{u}) = \mathbf{x}_i'\mathbf{u}$
Polynomial	$K(\mathbf{x}_i, \mathbf{u}) = (\gamma(\mathbf{x}_i'\mathbf{u}) + \mathbf{1})^{degree}$
Hyperbolic tangent (sigmoid)	$K(\mathbf{x}_i, \mathbf{u}) = tanh(\gamma(\mathbf{x}_i'\mathbf{u}) + \mathbf{1})$
Radial basis function (RBF)	$K(\mathbf{x}_i, \mathbf{u}) = \exp(-\gamma \parallel \mathbf{x}_i - \mathbf{u} \parallel^2)$

The SVM regression model minimises the following cost function:

$$C \sum_{i=1}^{n} L_\epsilon(y_i - f(x_i)) + \sum_{j=1}^{P} {\beta_j}^2 \tag{4.13}$$

where $L_\epsilon(.)$ is the ϵ-insensitive function and C is the cost penalty. The $L_\epsilon(.)$ function nullifies the residuals within the ϵ-insensitive tube. The C value helps adjust the model complexity by penalising large residuals. When C is large, the effect of residuals is amplified. Since the value of C has a large influence on the model performance, it is considered as an important hyperparameter.

For details on SVM algorithm and its hyperparameters, readers may refer [46,60,149]. The hyperparameters of SVM considered for optimisation in this research are the *kernel function*, ϵ, C *and* γ.

4.3.2.3 Gradient boosted trees (GBT)

GBT is an ensemble of decision tree (DT) algorithm. This section starts with the exploration of DT algorithm and later proceeds to boosted ensembles.

Starting from a root node, the DT algorithm generates a set of 'if-then-else' rules at each decision nodes below, until the tree terminates at the leaf nodes. Figure 53 represents these nodes within two different trees. One of the most competent techniques for constructing a DT, the classification and regression trees (CART) methodology [150] is adopted in this research.

The mathematical formulation given further is from [50]. The DT regression algorithm searches a training set S to find a predictor and split-value that partitions the data samples into two groups (S_1 and S_2) at the first decision node (root node). This is performed based on the minimisation of a *splitting criterion*. For example, using the sum of squared errors (SSE) as a criterion, the split is achieved by minimising:

$$SSE = \sum_{i \in S_1}(y_i - y_1)^2 + \sum_{i \in S_2}(y_i - y_2)^2 \tag{4.14}$$

where y_1 and y_2 are the averages of the training-set responses within the respective groups. The predictor with the lowest SSE splits the node in consideration into two new nodes below. This method is repeated in search of the respective predictor and split-value within each of the groups.

This recursive partitioning method of data splitting grows the tree until the number of samples in each group falls below a pre-set threshold, represented by the hyperparameter *minimum samples to split*. The number of samples required to be at the leaf node could also be pre-set as a hyperparameter – *minimum samples in leaf*. The *maximum depth of a tree* is the distance from its root node to the farthest leaf node. The hyperparameters, *minimum samples to split, minimum samples in leaf* or *maximum depth of a tree* may be used to tune the size of the DT. Increase in size increases the complexity of decision rules and may result in over-fitting. So, it is important to find the optimal tree for the given training-set samples.

The set of decision rules in the DT algorithm are highly interpretable and easy to implement, making it a favoured machine learning algorithm. DT enables feature selection through the measurement of relative importance of predictor variables in the predictions. Since this is inherent in the algorithm, it is not necessary to perform feature selection prior to training. This could be achieved by aggregating the reduction in *SSE* (or other splitting criterion used) for the training set over each predictor variable. Intuitively, the predictor variables being split in the upper nodes of the tree or those used multiple times are inferred to have more influence on the predictions. DT algorithm can handle heterogeneous data, and hence data scaling is not necessary. In addition, the DT algorithm is capable of handling missing data, although a large number of those would result in biased selection of predictor variables.

DT based models have high variance and a small change in the training data could result in a different set of splits. Ensembles are particularly useful in solving this problem. Boosting ensembles based on the ‘gradient boosting machines' developed by Friedman [52], follow the principle: given a loss function (such as least squares) and a weak learner (a trained base model with poor predictive performance), the algorithm seeks to find an additive model that minimises the loss function. DT base models are good candidates for boosting since they can be easily generated, tuned

(such as by changing the depth) and added sequentially. GBT algorithm which is an ensemble of DT regression base models, is discussed further.

A simple representation of a GBT ensemble made of two weighted decision tree base models is given in Figure 53. In GBT, additive models of the following form are considered:

$$F(x) = \sum_{m=1}^{M} \gamma_m h_m(x) \tag{4.15}$$

where $h_m(x)$ represents the DT base models of fixed size and γ_m the weight parameter. The models are built in a forward stage-wise fashion such that the model at the m^{th} stage:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{4.16}$$

Given the model fit $F_{m-1}(x_i)$ on n samples of the training set, $\gamma_m h_m(x)$ is obtained by minimising the loss function:

$$\sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma_m h_m(x)) \tag{4.17}$$

using steepest descent [52]. Loss functions such as least squares (LS), least absolute deviation (LAD), huber or quantile may be used.

The hyperparameters of GBT algorithm considered in the research are the *maximum depth of a tree*, the *minimum samples to split*, the *number of base models* and the *loss function*.

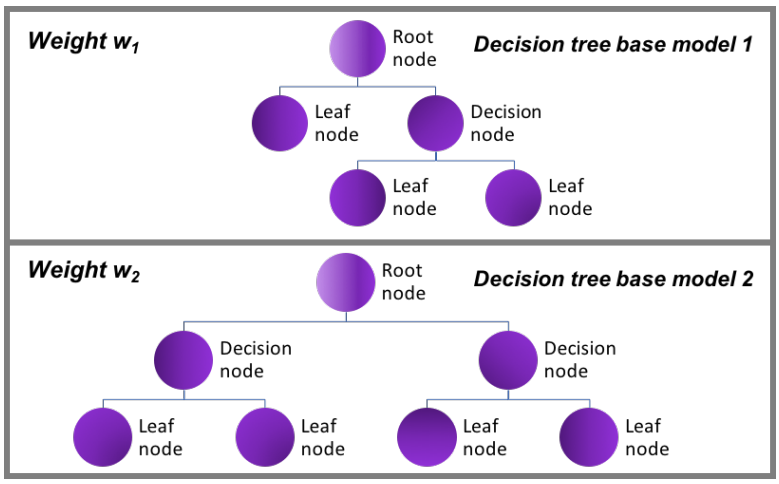

Figure 53: Ensemble of two weighted decision trees base models used in GBT

4.3.3 Hyperparameter optimisation

Methods of HPO are categorised into **heuristic** and **metaheuristic** [85]. Heuristic methods such as: 1) manual-search [60,86–89] enables simple computation, but at the cost of a limited search space, rendering good hyperparameter values that are not replicable by another modeller [81]; 2) grid-search is widely adopted [83,90–93] for its extensive search space and replicability, although affected by the curse of dimensionality making it a computationally slow method [81]. Metaheuristic methods such as evolutionary/genetic algorithm [94,95] and swarm algorithms [39–43] have been implemented. However, these are computationally demanding [101] and have bigger storage requirement [98]. The implementation of metaheuristic methods requires more code complexity and demands domain expertise for initialisation of respective parameters [81]. These were reviewed in detail in Chapter 3 Section 3.3.3.

Since the data-driven modelling developed in this chapter targets DR applications, computational efficiency is prioritised while choosing the HPO method. A large number of real-life optimisation problems can be tackled using heuristic methods [85]. However, those being widely applied for HPO in ML-based building load estimation perform on the extremes. While manual-search is not replicable, grid-

search is computationally inefficient. As an improvement over these, Bergstra and Bengio [81] developed a heuristic method of HPO called random-search.

In terms of performance on high-dimensional hyperparameters, random-search HPO is proven to be superior to grid-search based on its application on different ML algorithms and datasets. Unlike grid-search, all combinations of hyperparameter values in the search space are not tried out. Rather a fixed number of sets are randomly sampled, hence making it computationally efficient. The sampling size could be altered based on the available computational resources, giving more control to the modeller [81]. Considering these advantages, random-search HPO is implemented for model calibration in the presented study. The mathematical description for this method is provided in the section below.

4.3.3.1 **Random-search HPO explained**

In random-search HPO, trial hyperparameter values are numbers or keyword predetermined by the modeller. The numbers are either given as distributions or listed as integers depending on the type of the hyperparameter. The modeller may choose a broad range of numbers. Trial values form hyperparameter sets that constitute a hyperparameter search space of dimension d i.e. $\boldsymbol{\Lambda} = \{\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)} \dots \boldsymbol{\lambda}^{(\mathrm{d})}\}$.

A learning algorithm using hyperparameter set $\boldsymbol{\lambda}$, denoted as $A_{\boldsymbol{\lambda}}$, finds a function:

$$f(x) = A_{\boldsymbol{\lambda}}(\boldsymbol{X}^{(\mathrm{train})}) \tag{4.18}$$

for a training set $\boldsymbol{X}^{(\mathrm{train})}$ drawn out of independent and identically distributed (i.i.d) samples x from a distribution G_x such that it minimises some expected loss $\mathcal{L}(x; f(x))$. We refer to $A_{\boldsymbol{\lambda}}$ as a prototype model since it does not represent the final calibrated model. The good hyperparameter set $\hat{\boldsymbol{\lambda}}$ is selected from the hyperparameter search space $\boldsymbol{\Lambda} = \{\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)} \dots \boldsymbol{\lambda}^{(\mathrm{d})}\}$ by minimising the mean loss. This is performed using cross-validation over a validation set $\boldsymbol{X}^{(\mathrm{valid})}$ that consists of i.i.d samples drawn from G_x. The cross-validation is unbiased since the samples in $\boldsymbol{X}^{(\mathrm{valid})}$ are independent of those used by $A_{\boldsymbol{\lambda}}$. The optimisation addressed here is represented in the equations 4-5.

$$\hat{\boldsymbol{\lambda}} \approx \underset{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}}{\operatorname{argmin}} \quad \underset{x \in \boldsymbol{\mathcal{X}}^{(\text{valid})}}{\text{mean}} \quad [\mathcal{L}(x; A_{\boldsymbol{\lambda}}(\boldsymbol{\mathcal{X}}^{(\text{train})}))] \tag{4.19}$$

$$\approx \underset{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}}{\operatorname{argmin}} \quad \boldsymbol{\Psi}(\boldsymbol{\lambda}) \tag{4.20}$$

Here, Ψ is the hyperparameter response function and HPO is the minimisation of $\boldsymbol{\Psi}(\boldsymbol{\lambda})$ over $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$. The learning algorithm $A_{\hat{\boldsymbol{\lambda}}}$ with the good hyperparameter set $\hat{\boldsymbol{\lambda}}$ is referred to as the calibrated forecast model for the given dataset [81].

In grid-search, if k is the number of hyperparameters to be optimised and n is the number of trial values of each hyperparameter, dimension of the search space hence formed would be $d = n^k$. In random-search, a fixed number of values referred to as the sampling size s, are randomly sampled from the trial values to form hyperparameter sets. Formation of a search space from trial values of the SVM hyperparameters (kernel function, ε, C and γ) using random-search method with $s = 3$, is visualised in Figure 54. In this example, trial values are sampled with replacement, and is noticeable for hyperparameter C. The dimension of the search space obtained in random-search is the same as its sampling size, i.e. $d = s$.

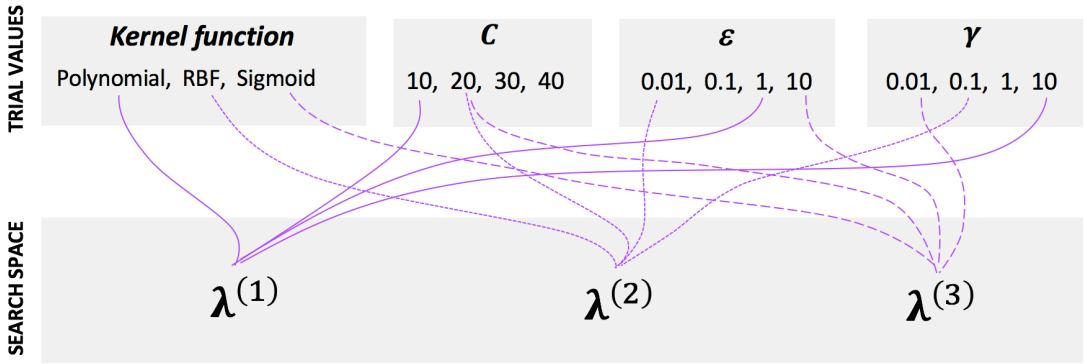

Figure 54: Search space obtained by random sampling of trial hyperparameters values of SVM using random-search method (sampling size=3)

4.3.4 Model calibration, selection and evaluation

The ML algorithms considered in this research are supervised regression approaches of ANN, SVM and GBT. The architectures of these algorithms along with their hyperparameters were discussed in detail in Section 4.3.2. Random-search HPO methodology adopted in the study was explained in the previous section. In this section, the calibration of ML algorithms using random-search HPO, selection of the best performing models and their evaluation are elaborated. These are then demonstrated on the office building dataset.

Hyperparameters optimised for each of the learning algorithms are: 1) ANN: the *hidden layer size* (number of hidden layers and numbers of hidden layer neurons), the *activation function*, the *regularisation term* and the *optimisation technique*; 2) SVM: the *kernel function*, ε, C *and* γ; and 3) GBT: the *maximum depth of a tree*, the *minimum samples to split*, the *number of base models* and the *loss function*.

The hyperparameter values maybe numbers or keywords. Integer trial values are given as discrete distributions and real values as continuous distributions. A discrete distribution within the bounds a and b is represented as DD(a,b). For example, the trial values for the *maximum depth of a tree* of the GBT algorithm is listed as DD(1,100), namely the integers 1–100. A continuous distribution is represented as CD(a,b). For example, trial values for ε in the SVM is a continuous distribution, CD(0.01,10). Trial values of some of the numerical hyperparameters such as the *number of base models* of GBT and keyword hyperparameters such as the *activation function* of ANN are simply listed in square brackets '[ ]'.

A sampling size of 50 is considered for random sampling from the trial hyperparameter values of each ML algorithm. This helps build a search space of dimension 50, which is the number of hyperparameter sets formed by random sampling from trial values. Sampling is performed with replacement from the trial hyperparameter values, as exemplified earlier in Figure 54. Since the sampling size is small, multiple stages of random-search HPO could be easily performed using the

available computational resources. Here, fine calibration of each ML algorithm is discussed using two stages of random-search HPO. This is represented with a block diagram in Figure 55.

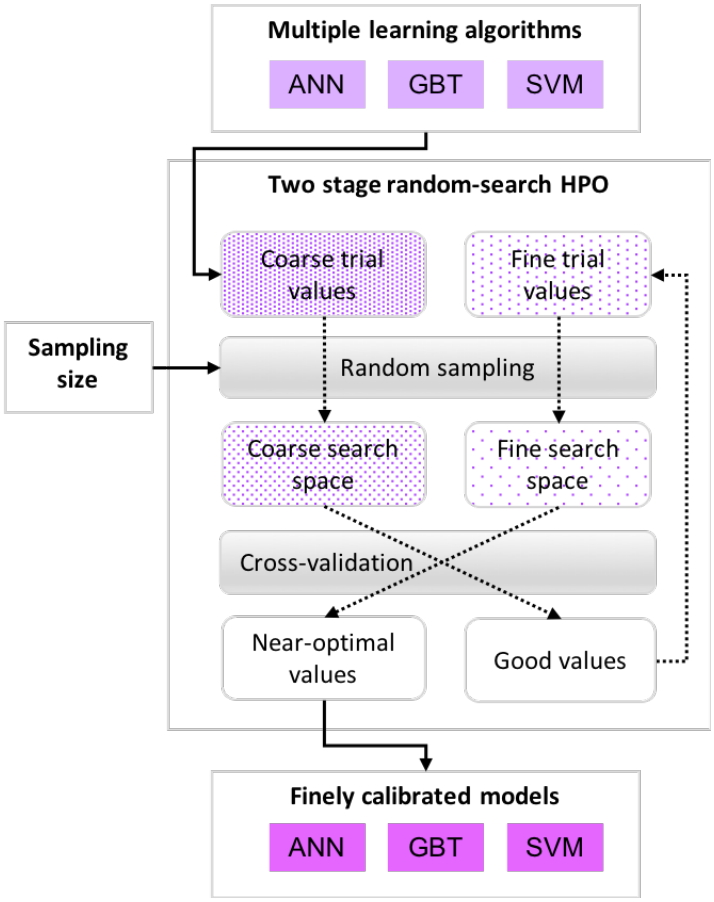

Figure 55: Fine calibration using two stage random-search HPO on multiple ML algorithms

It is difficult to visualise multistage HPO for high-dimensional hyperparameters. With the help of a 2D contour plot, Figure 56 represents the two stage HPO of two hyperparameters (γ and C) of the SVM algorithm.

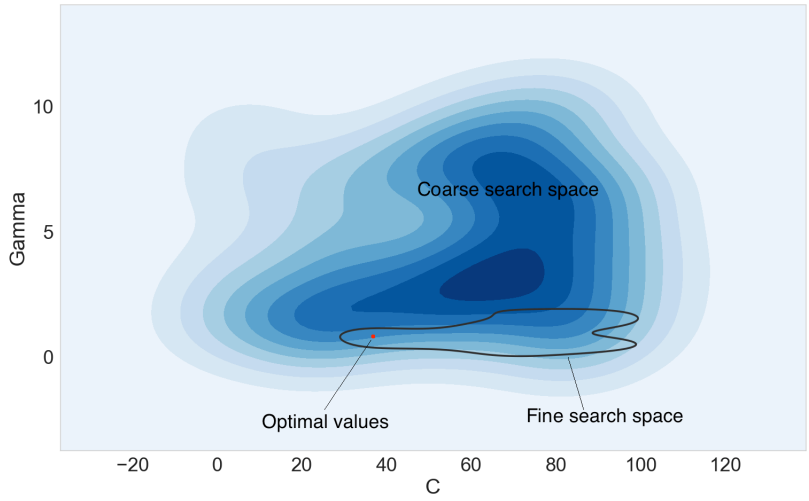

Figure 56: Two stage HPO on γ (gamma) and C hyperparameters of SVM to identify the near-optimal values

Fine calibration, model selection and evaluation of each ML algorithm are performed based on the following steps.

- Trial hyperparameter values are predetermined for the first stage of random-search HPO. A broad range of trial values are used and these are referred to as coarse trial values. Using the pre-set sampling size of 50, a coarse search space of dimension 50 is randomly sampled from the coarse trial values.
- Hyperparameter sets in the coarse search space are used in the ML algorithm to form unique prototype models. As part of cross-validation, each of these prototype models are trained and tested on randomly sampled training-testing sets of the development set. A prototype model showing the best predictive performance using a statistical metric is considered as the calibrated model of the learning algorithm on the building dataset.
- During cross-validation, predictive performance is measured on the testing sets using the mean absolute error (MAE) metric. The MAE of prototype models on every training-testing set (in the development set) is recorded and their mean is calculated. The lowest average MAE is referred to as the development set error (DSE) of the coarsely calibrated model, with a reasonably good set of

hyperparameter values. This coarsely calibrated model is validated on the training-testing sets in the evaluation set. The R^2 (in %) accuracy metric is used to measure predictive performance on the testing sets in the evaluation set and their average is referred to as the evaluation set accuracy (ESA). The R^2 (in %) accuracy metric is used for better interpretation of the model performance.

- As part of the second stage of HPO, fine trial values in the proximity of the good set of values identified in the previous step are randomly sampled to build a fine search space. A sampling size of 50 is used in this stage as well. New prototype models are formed based on the hyperparameter sets in the fine search space.
- Cross-validation of these prototype models is performed again on the development set and the model with lowest mean error (recorded as DSE) is expected to be finely calibrated on the building dataset. Based on this, better (referred to as *near-optimal* henceforth) hyperparameter values of the ML algorithm are identified.
- Finely calibrated model of the learning algorithm is validated again on the training-testing sets of the evaluation set to obtain ESA.

The above-mentioned steps are repeated for all the three ML algorithms. The finely calibrated models of each ML algorithm are compared based on their predictive performances on the evaluation set. The best model is subsequently selected and its ESA is recorded as the performance benchmark for future reference. This best model is then deployed in DR operation for the respective building. Based on this methodology, model calibration, selection and evaluation are performed on the office building dataset for a day-ahead and a week-ahead load estimation model, in the next section.

4.3.4.1 **Demonstration on the office building dataset**

The response variable values consist of half-hourly meter data measured for one year. Predictors such as time-of-day, day-of-week, daily degree days from previous day and daily mean/maximum/minimum ambient dry-bulb temperatures from previous day are used to develop the day-ahead model. Similarly, predictors such

as time-of-day, day-of-week, weekly degree days from previous week and weekly mean/maximum/minimum ambient dry-bulb temperatures from previous week are used to develop the week-ahead models. These predictors belong to the class I category discussed in Section 4.2.4. The testing set size is 1 day for day-ahead and 1 week for week-ahead models. The training set sizes are taken as 13 weeks for both the day-ahead and week-ahead models based on the experimentation performed in Section 4.3.1.5. The training-testing sets are selected on a sliding window basis from the annual dataset such that the windows slide forward in steps of 1 day for the day-ahead model and in steps of 1 week for the week-ahead model. From these, 40 training-testing sets are randomly sampled and equally distributed into the development set and evaluation set, for the respective models. For each ML algorithm, the coarse trial values used to build the coarse search space (of dimension 50), the good set of values obtained after calibration and the model performances from the first stage of random-search HPO are listed in Table 6 (day-ahead model) and Table 7 (week-ahead model).

Table 6: First stage of model calibration using coarse search space for the day-ahead model

ML algorithm	Hyperparameters	Coarse trial values	Good values	DSE	ESA(%)
ANN	*Size of hidden layer*	[(14,),(14,14),(14,14,14)]*	(14,14,14)	0.22	82.9
	Activation function	[Logistic, Tanh, Relu]	Logistic		
	Optimisation technique	[LBFGS, SGD]	LBFGS		
	Regularisation	CD(0.0001,0.1)	0.08		
SVM	*Kernel function*	[Polynomial, RBF, Sigmoid]	RBF	0.19	88.4
	C	CD(1,100)	30.1		
	ε	CD(0.01,10)	0.65		
	γ	CD(0.01,10)	0.71		
GBT	*Number of base models*	[100,250,500,750,1000]	250	0.17	90.8
	Loss function	[LS, LAD, Huber]	LAD		
	Max. depth of a tree	DD(1,100)	24		
	Min. samples to split	DD(1,100)	10		

* For the size of hidden layer in ANN, (14,) represents 1 hidden layer with 14 neurons, (14,14) represents 2 hidden layers with 14 neurons in each layer and so on.

Table 7: First stage of model calibration using coarse search space for the week-ahead model

ML algorithm	Hyperparameters	Coarse trial values	Good values	DSE	ESA(%)
ANN	*Size of hidden layer*	[(14,),(14,14),(14,14,14)]	(14,14)	0.27	79.9
	Activation function	[Logistic, Tanh, Relu]	Relu		
	Optimisation technique	[LBFGS, SGD]	LBFGS		
	Regularisation	CD(0.0001,0.1)	0.01		
SVM	*Kernel function*	[Polynomial, RBF, Sigmoid]	RBF	0.22	84.1
	C	CD(1,100)	24.4		
	ε	CD(0.01,10)	0.68		
	γ	CD(0.01,10)	0.30		
GBT	*Number of base models*	[100,250,500,750,1000]	500	0.19	88.8
	Loss function	[LS, LAD, Huber]	LAD		
	Max. depth of a tree	DD(1,100)	38		
	Min. samples to split	DD(1,100)	16		

As can be seen, a broad range of trial hyperparameter values have been used (wherever possible). The coarse search space that contains 50 hyperparameter sets formed by random selection of trial values has not been listed. The GBT algorithm shows higher evaluation set accuracy over SVM and ANN for the day-ahead as well as the week-ahead models. However, no model selection based on predictive performance is conducted at this stage. Rather, the identified good set of hyperparameter values help build fine trial values that are used in the second stage of random-search HPO.

The fine search space that contains 50 hyperparameter sets formed by random selection of fine trial values for the second HPO stage is not listed. The good hyperparameter values, development set error (DSE) and evaluation set accuracy (ESA) in this stage are given in Table 8 (day-ahead model) and Table 9 (week-ahead model).

Table 8: Second stage of model calibration using fine search space for the day-ahead model

ML algorithm	Hyperparameters	Fine trial values	Near-optimal values	DSE	ESA (%)
ANN	*Size of hidden layer*	[(14,14,14)]	(14,14,14)	0.21	83.2
	Activation function	[Relu]	Relu		
	Optimisation algorithm	[LBFGS]	LBFGS		
	Regularisation	CD(0.01,0.1)	0.024		
SVM	*Kernel function*	[RBF]	RBF	0.17	89.1
	C	CD(25,75)	28.5		
	ε	CD(0.1,1)	0.22		
	γ	CD(0.5,2)	1.2		
GBT	*Number of base models*	[150,200,250,300,350]	300	0.16	91.3
	Loss function	[LAD]	LAD		
	Max. depth of a tree	DD(15,40)	28		
	Min. samples to split	DD(1,20)	13		

Table 9: Second stage of model calibration using fine search space for the week-ahead model

ML algorithm	Hyperparameters	Fine trial values	Near-optimal values	DSE	ESA (%)
ANN	*Size of hidden layer*	[(14,14,14)]	(14,14,14)	0.24	82.1
	Activation function	[Logistic]	Logistic		
	Optimisation algorithm	[LBFGS]	LBFGS		
	Regularisation	CD(0.005,0.1)	0.03		
SVM	*Kernel function*	[RBF]	RBF	0.20	86.2
	C	CD(25,75)	28.5		
	ε	CD(0.1,1)	0.2		
	γ	CD(0.5,2)	1.8		
GBT	*Number of base models*	[150,200,250,300,350]	300	0.18	89.8
	Loss function	[LAD]	Huber		
	Max. depth of a tree	DD(15,40)	70		
	Min. samples to split	DD(1,20)	51		

The finely calibrated day-ahead models are evaluated first. The ANN has an ESA of 83.2%, whereas the SVM has an ESA of 89.1%. The forecast performance of SVM model is better than that of ANN model. The GBT is finely calibrated with an ESA

of 91.3%. The GBT model shows improvement in the predictive performance over SVM. The finely calibrated week-ahead models also show similar trends in performance as that of the day-ahead models, with ESA of 82.1% for ANN, 86.2% for SVM and 89.8% for GBT.

Overall results show that, when compared with the outcomes of the first stage of HPO (coarse) for all three ML algorithms, there is improvement in respective predictive performances after the second stage (fine). This validates the fact that multiple stages of HPO help in finely calibrating ML based forecast models and hence improving their predictive performance. The use of random-search HPO method, enables fine calibration through efficient allocation of available computational resources. The only user inputs during random-search HPO are the ranges of trial hyperparameter values and the sampling size, that do not require much domain expertise.

The total runtime for two stages of HPO for ANN was ~1.8 hours, SVM was ~ 2.2 hours and GBT was ~1.3 hours. These are not considered as a criterion for model selection since model calibration is an offline process repeated only when the predictive performance deteriorates over time (discussed further in the next section). Hence model selection is performed purely based on their predictive performance in the evaluation set. The GBT based models prove to be better in terms of the accuracies achieved.

4.4 Model deployment

The three stages of data-driven modelling presented in this chapter, namely data pre-processing, model development and model deployment, and the associated tasks within every stage are summarised in Figure 57. The first two stages in this data-driven pipeline discussed in the preceding sections are oriented towards applications where large number of consumer buildings are involved. While data-driven models are inherently adaptable to new data, the specific focus of the

modelling has been on computational efficiency and replicability. In this section, aspects related to deployment of the data-driven building load estimation model in real-world scenarios are discussed.

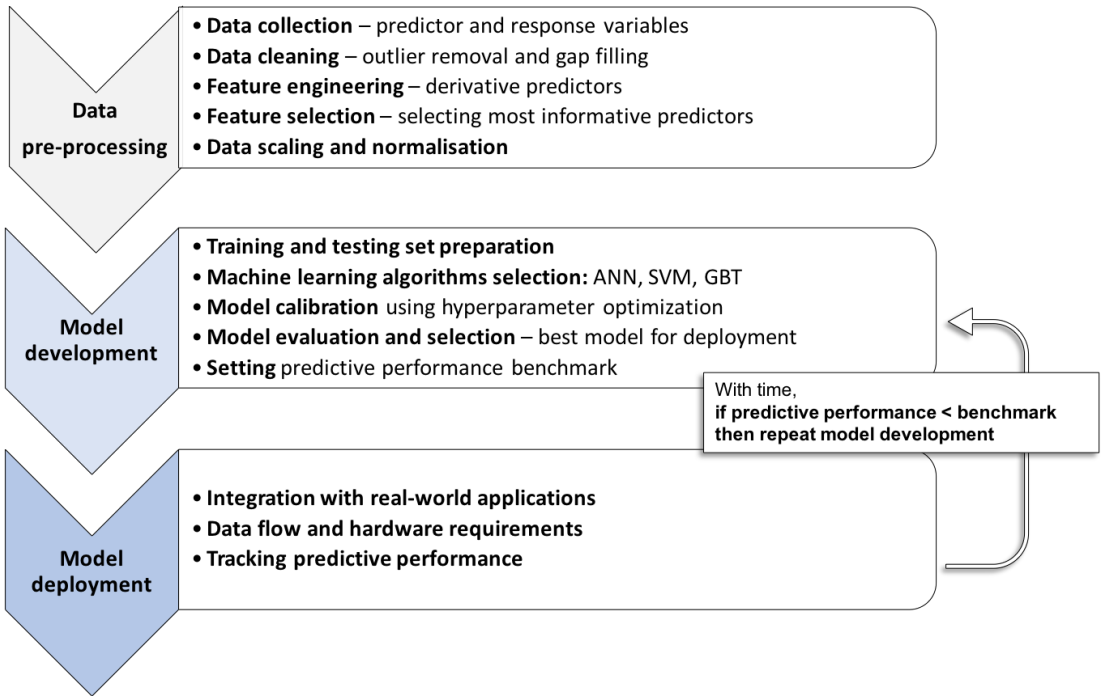

Figure 57: Data-driven pipeline for building load estimation modelling

4.4.1 Integration with real-world applications

The modelling approach developed in this chapter is applied for building load estimation in DR operational tasks such as capacity scheduling, performance evaluation (in Chapter 5) and reliable operation (in Chapter 6). Each of these tasks are designed around ongoing DR program frameworks and the energy asset used to deliver the DR. These are accounted as part of the data-driven model development in Chapter 5 and Chapter 6.

DR programs define specific requirements in terms of the estimation horizon (windows) and recurrence of these estimates. For example, the frequency control by demand management (FCDM) program in the UK mandates week-ahead windows of the load curtailment capability recurring every week. Further, an hour-ahead

adjustment is also allowed in the FCDM contracts. This would mean that the building participating in this program will have a week-ahead as well as an hour-ahead load estimation model deployed in the DR operations. It is important that the data-driven model development considers such specific requirements for each building. Since the DR program frameworks are liable to changes, it is also essential to repeat the model development based on the new specifications and re-deploy the model in operations.

The type of energy asset that delivers DR, such as flexible loads, standby generators and storage systems, also influences the data-driven model development. For load curtailment DR programs, the requirement is to estimate the curtailable load. If the consumer energy assets such as standby generators or storage systems are used, whole building load could potentially be curtailed. In these cases, whole building load is the curtailable load. If only flexible loads are available, the amount of flexibility needs to be estimated. When sub-meter data are not available, the whole building load estimation models could be used in conjunction with available information on the flexibility in the building. If sub-meter data are available, load specific ML models could be developed.

4.4.2 Data flow and hardware requirements

After model development, the best supervised ML model is deployed in DR operations. Selected day-ahead and week-ahead load estimation models for the demonstrated office building dataset use class I predictors such as, 1) temporal predictors: time-of-day, day-of-week; and 2) weather predictors: daily/weekly mean/extrema of ambient dry-bulb temperature lag values, daily/weekly degree days lag values. The response variable is the half-hourly meter data that can be collected from the building site in real-time and stored in a database. Temporal predictors can be derived from timestamps of these meter data. Representative weather data can be collected from on-site weather measurements (which is the most preferable) or the nearest meteorological stations and stored in the same database as the meter data. If class II predictors were used in the deployed model,

weather forecasts will have to be accessed from reliable sources. However, our analysis has shown that the temperature forecasts and their derivatives (mean, degree days, etc.) did not substantially improve the predictive performance of the models, beyond what was already achieved using their lag values. This also helps avoid uncertainty associated with including weather forecasts in the model.

A common database consisting of the temporal and weather predictors (and building predictors if used) may be stored in a web-based or a local server. Since the models are trained using 13 weeks of data prior to the period for which the building loads are being forecasted (day-ahead or week-ahead), continuous data accessibility from this database needs to be ensured. Based on the experimentation in Section 4.3.1.6, data gaps between the training and testing sets are not recommended.

Data-driven modelling performed in this research has focussed on computational efficiency. The models were developed in the open-source Python programming language environment using the Scikit-Learn ML module [46] on an Intel Core i5, 2.5 GHz, 64-bit, quad-core computer with 8 GB RAM. These resources are easily available in the academic and industrial establishments. In the whole modelling process, model calibration using HPO is the most computationally intensive process. The ML module used makes use of multiple CPU cores for parallel processing to speed up this process. In the deployment context, model calibration need not be performed repeatedly and the selected best model can produce continuous forecasts in an even lower configuration computer.

4.4.3 Tracking the predictive performance

While deployment of the best data-driven model is expected to deliver the best estimates, it may so happen that the predictive performance deteriorates over time. In such instances, it is advisable to refer back at the benchmark performance recorded during model selection. If predictive performance of the deployed model has indeed deteriorated, the model development stage is repeated and the 'updated' best model is deployed again. This is also represented as part of the data-driven

pipeline in Figure 57. It can be a good practice to keep logs of different versions of the model, so that the updates can be traced back. Performance tracking is particularly important for enabling fully automated continuous forecasting in DR operations and similar applications.

4.5 Conclusions

As larger datasets become increasingly available, data-driven modelling based on ML makes great sense. Progress in the field of empirical ML enables faster, easier and cost-effective development of high performance models. This has opened a gateway to bypass the complexities in the development of detailed physics-based or hybrid models.

A data-driven modelling approach for building load estimation was developed and demonstrated in this chapter. The modelling was performed based on specific design criteria such as reliability, computational efficiency, potential for large scale deployment and automation, for application in DR operations. These design criteria were considered in every stages of the modelling process. Numerical experimentations and diagnostics were performed using an office building dataset.

The first stage of data pre-processing involves data collection, data cleaning, feature engineering, feature selection and data scaling. A non-parametric method, one-class SVM was proposed for outlier removal as part of data cleaning process. This unsupervised ML algorithm learns the relevant data without the need for training examples and provides sufficient control for the modeller. Feature engineering was performed to derive more informative predictors from the collected meter data timestamps and weather variables. Filter methods were proposed for feature selection since they are computationally efficient and reduces the iterations during model development.

As part of the second stage i.e. model development, ML algorithms and their hyperparameters were described. Although only 3 ML algorithms were used

(artificial neural networks, support vector machines and gradient boosted trees) the open design encourages the use of more ML algorithms. Random-search hyperparameter optimisation (HPO) method for model calibration was proposed for its computational efficiency and reliability. Following this, model calibration, selection and evaluation were discussed and demonstrated for day-ahead and week-ahead building load estimation models. The GBT models were observed to outperform the ANN and SVM models. The multistage HPO using random-search method was demonstrated to utilise the available computational resources and improve the model performance in stages.

Aspects related to the third stage, i.e. model deployment, were also explored. This included discussions on the integration with real-world applications such as DR operations. Data flow and hardware requirements were also summarised. The need for tracking the model's predictive performance based on a benchmark defined during its development was highlighted. In a continuous operational scenario, this provides the trigger for model re-calibration and re-deployment, hence enabling increased automation.

With the availability of smart grid data and off-the-shelf ML software packages, the developed approach could be easily replicated on large number of buildings to build data-driven models using minimal computational resources. Such minimalistic models may also be deployed in large-scale through connected devices on consumer sites supported by the internet-of-things (IoT) platforms.

The data-driven modelling developed, demonstrated and evaluated in this chapter is used for applications in specific DR operational tasks in the upcoming two chapters. These are: capacity scheduling, performance evaluation in Chapter 5 and reliable operation in Chapter 6.

Chapter 5

Data-driven capacity scheduling and performance evaluation for DR from large consumer buildings with load curtailment capability

5.1 Introduction

Demand response (DR) mechanism and the related aspects were introduced in Chapter 2. This chapter addresses two operational tasks in the context of incentive-based DR programs for large building consumers where load estimation is necessary: *capacity scheduling* and *performance evaluation*. The data-driven building load estimation approach developed and demonstrated in the previous chapter is applied in these DR operational tasks. Performances of the data-driven models are then compared with that of the conventional models used in ongoing incentive-based DR programs in different electricity markets. The presented study validates the utility of data-driven modelling in real-world DR applications.

5.1.1 Background

Large consumer buildings with energy assets such as flexible loads, standby generators or storage systems can deliver load curtailment for grid balancing. This is because of the interaction of these energy assets with the whole building load. For instance, when standby generators or storage systems start backing up the whole building load, the grid would see it as full load curtailment. In the incentive-based DR programs, consumers are usually rewarded for the availability and utilisation of their load curtailment capacity.

For large consumer buildings participating in incentive-based DR programs, *capacity scheduling* is the task of estimating their load curtailment availability and informing the program operator. As part of the planning, prior to the DR event, program operators mandate DR capacity scheduling from the consumers for pre-contracted windows. Based on this information, an availability payment is often made to the participant, regardless of a DR event occurring in that window. However, depending on the model used, it is possible that the estimated load curtailment and hence the scheduled capacity, is not accurate. One of the uncertainties faced by DR program operators is regarding the actual capacity that would be delivered during the grid event with respect to the scheduled capacity [151]. This uncertainty could be minimised by adopting reliable models for accurate building load estimation, contributing towards a better capacity scheduling practice in DR operations.

Once a DR event has occurred in the pre-contracted window, the program operator conducts a *performance evaluation* of the building consumer by measuring the actual metered load (reflecting the load curtailment) with reference to a baseline that would have occurred in the absence of the DR event [105]. Based on the difference between the baseline load and the metered load, a utilisation payment is often made to the consumer. Underestimation of the baseline load would result in lower utilisation payment for the consumer, while overestimation would result in a loss for the program operator. This reinstates the need for a reliable building load estimation model to be used in performance evaluation as part of the DR operations.

A representative DR event induced building load curtailment and commonly used DR program parameters are shown in Figure 58, with the capacity scheduling and performance evaluation tasks in perspective.

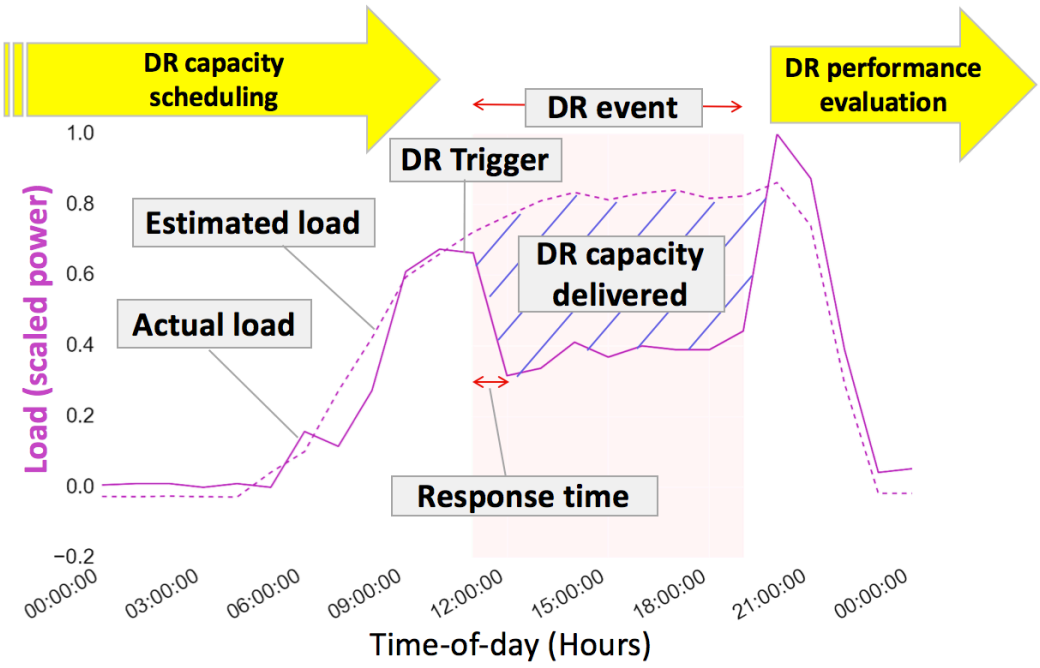

Figure 58: DR capacity scheduling and performance evaluation in perspective

In many of the ongoing incentive-based DR programs, there are no operator prescribed models for capacity scheduling. Hence, consumers or DR aggregators commit their own estimates of the capacity availability for a given window to the DR program operator. Against this, there are many conventional models adopted by the program operators for building baseline load estimation towards performance evaluation, however none of them are standardised across the operator regions (or electricity markets). These models include: day averaging and statistical regression [105,152]. Day averaging models consider 'x days out of the y like days' (which are separately considered for weekdays and weekends/holidays) from the near past, based on a criterion, and average the energy consumption for the event interval from those x days. A day averaging with event adjustment model considers a correction factor to these estimates prior to the event interval. Studies have shown that regression models perform better than the day averaging models [153]. However, polynomial representations of the regression variables are required to adapt the regression models to non-linear data such as the building load, making them difficult to replicate on large number of buildings. Availability of smart meter

data enable the development of data-driven models for building load estimation using machine learning (ML) algorithms. Many ML algorithms are capable of learning the non-linear relationships between predictor variables and the building loads being estimated. These algorithms continue learning from the new incoming data, making them more adaptable to deployment and operational scenarios. A supervised ML based building load estimation modelling approach was developed and demonstrated in the Chapter 4.

Few previous studies have explored the use of data-driven modelling for DR capacity estimation in large consumer buildings. Nghiem and Jones [102] developed a Gaussian processes (GP) based supervised regression ML model for predicting the load response DR behaviour of commercial buildings using DR signals and weather variables as predictors. Jung et al. [103] estimated the available flexible DR capacity in two large buildings based on a data-driven model using data from building variables such as temperature/humidity/light sensors, carbon dioxide sensors, passive infrared sensors and smart plug power meters; the model is claimed to be better than the conventional manual audit processes used to estimate DR capacity.

Some studies have also used data-driven modelling for building baseline load estimation as part of DR operations. Behl et al. [104] demonstrated the superior performance of decision trees (DT) based data-driven model for baseline load estimation on 8 large buildings using building and weather variables. Chen et al. [105] proposed an SVM regression based data-driven model for baseline load estimation and demonstrated this on 4 office buildings; the baseline is used to evaluate DR performance for building load curtailment and is shown to perform significantly better than the conventional models based on day averaging and polynomial regression. Park et al. [106] showed that a data-driven model based on unsupervised ML techniques has lower errors in comparison to conventional day-matching methods for baseline load estimation.

5.1.2 Contributions

Capacity scheduling and performance evaluation are two important operational tasks in incentive-based DR programs. In order to meet the increasing demand for DR in the grid, higher participation of large consumer buildings with energy assets is to be ensured. Parallel to this, reliable, computationally efficient, cost effective and deployable models need to be developed for performing DR tasks. Data-driven models using ML can be used for capacity scheduling and performance evaluation in large consumer buildings with load curtailment capability. Few studies have explored the benefits of data-driven models in these DR operational contexts. However, they have seldom focussed on the large-scale deployment of such models by making them computationally efficient; the presented study attempts to fill this gap.

In this Chapter, the supervised ML based load estimation modelling developed in Chapter 4 is applied on multiple large consumer buildings such as supermarket, laboratory, hotel, retail store, office and hospital, with focus on their potential deployment in DR tasks such as capacity scheduling and performance evaluation. These DR tasks are designed based on frameworks of ongoing incentive-based DR programs in the US and the UK electricity markets. Deployment related aspects such as duration of the load curtailment events in the DR programs are reflected in the development of data-driven models in this chapter. The performances of these models are compared with that of the conventional models used in the respective DR programs. The study shows that supervised ML models can outperform the conventional models in terms of the accuracy in building load estimation. Large-scale deployment of such data-driven models in DR operations can increase the reliability of capacity scheduling and performance evaluation, benefitting the grid as well as the building consumer.

5.1.3 Layout

The rest of the chapter is structured as follows. Section 5.2 elaborates on the framework of DR programs considered. Section 5.3 discusses the DR strategies for load curtailment in commercial buildings. Section 5.4 develops the data-driven model for building load estimation and discusses the conventional models adopted in the electricity markets where the DR programs are considered. Section 5.5 critically analyses the results and Section 5.6 concludes the chapter.

5.2 DR programs framework

Legacy grid balancing services such as frequency response, emergency reserves, capacity market mechanism, etc. operated by transmission system operators (TSOs) were primarily designed for centralised generators. Today's wholesale DR programs try to facilitate consumer participation in these services such that all technologies can compete in the market. For the purpose of this study, two representative incentive-based wholesale DR programs are considered: 1) *Frequency control by demand management (FCDM)* in the UK, and 2) *Emergency Response Service (ERS)* in Texas, US. It has to be noted that, these programs and their requirements may be modified in the future [154] and their selection in this study is for case building purposes only.

5.2.1 FCDM in the UK

FCDM is a wholesale DR program operated by National Grid (NG) in the UK, that helps minimise fall in system frequency through the interruption of large customers from the transmission system. A minimum individual or aggregated load curtailment capacity of 3 MW is required for participation in the FCDM program. Whenever the system frequency falls below the threshold level of 49.7 Hz, an equipment installed at the participant's site detects this and triggers the response. The participant is expected to deliver load curtailment within 2 seconds and sustain for a maximum duration of 30 minutes. Such events are expected to occur 10 to 30

times per year. An FCDM event and subsequent response of a participant load is represented in Figure 59. The FCDM program mandates week-ahead DR capacity scheduling at 30 minutes resolution. Based on this, the participants are paid an availability fee [155,156].

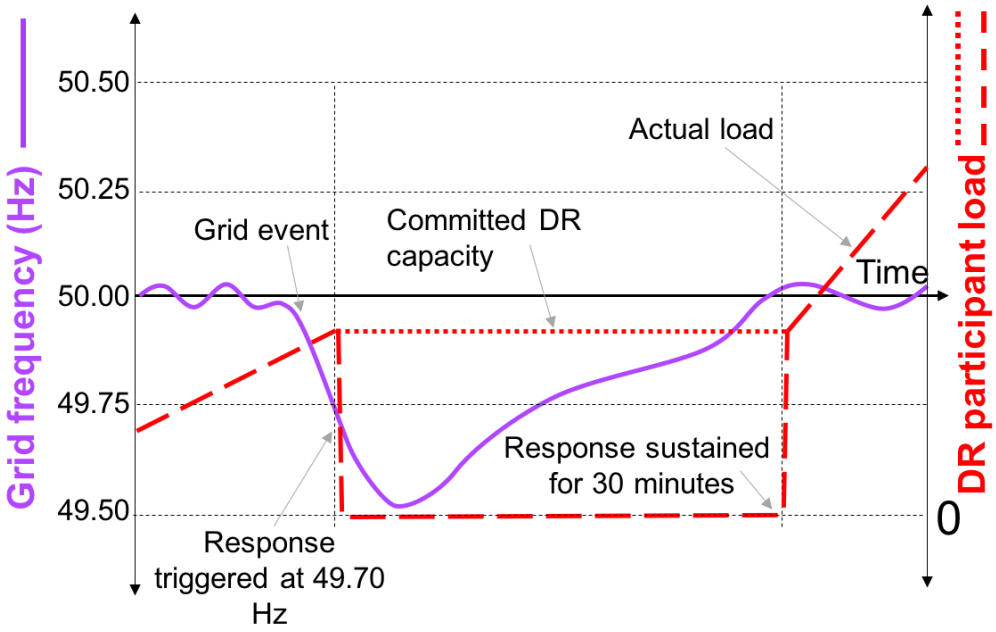

Figure 59: An FCDM event triggered load curtailment

5.2.2 ERS in the US

The electric reliability council of Texas (ERCOT) is the TSO in the state of Texas. ERCOT operates wholesale DR programs for consumers with eligible energy assets. The ERS operated by ERCOT is a DR program in which flexible loads and generators are selected for deployment as reserves during grid emergency events. Participating energy assets are classified as weather-sensitive or non-weather-sensitive and procured based on their response times that may vary between 10 to 30 minutes.

The ERS capacity procurement is conducted three times a year for different windows, listed in Table 10. Windows 1 to 5 are defined for weekdays (Monday-

Friday) and non-holidays only. ERCOT identifies the windows 1, 3 and 4 as appropriate for weather-sensitive assets. Participants with a minimum capacity of at least 500 kW may be procured for one or more windows, individually or through an aggregator [157]. The ERS load curtailment from a weather-sensitive flexible load may be visualised as shown in Figure 58. The performance evaluation is expected to be performed at 15 minutes resolution for 3 hours in each window or the duration of the DR event, whichever is smaller.

Table 10: Service delivery windows for emergency response service of ERCOT

Window	Duration
1	5 a.m. to 8 a.m.
2	8 a.m. to 1 p.m.
3	1 p.m. to 4 p.m.
4	4 p.m. to 7 p.m.
5	7 p.m. to 10 p.m.
6	All other hours

5.3 DR strategies

In large consumer buildings, different DR strategies for load curtailment may be adopted. Whether it is whole or partial load curtailment depends on the type of energy assets available in the respective buildings. These are elaborated below.

5.3.1 Whole building load curtailment

Standby generators are of interest in large consumer buildings for providing load curtailment in DR programs [158]. In this study, diesel rotary UPS (DRUPS) that combines standby diesel-generators with a flywheel UPS system providing instantaneous and continuous backup, is considered for whole building load curtailment. During grid interruption, stored kinetic energy in the flywheel is released into a motor-generator within the DRUPS unit, to sustain the consumer load for a few seconds, until the diesel engine takes over [159]. The connection

configuration of a DRUPS unit on a large consumer building site is represented in Figure 60.

DRUPS units enable fast transfer of whole building load from the grid during interruptions. These are seldom utilised for continuous emergency power since long duration grid outages are rare. Depending on the availability, these responsive energy assets add value to the grid through DR programs. The response time and possibility of whole building load curtailment enable commercial building consumers with DRUPS units to participate in the FCDM program. In the absence of DRUPS units, DR from commercial buildings would have been limited to curtailment of flexible loads and may not be eligible for FCDM due to the minimum capacity requirement of 3 MW for participation. Availability of the whole building load for curtailment results in higher revenue for the participant and this DR strategy has minimal interference with the building internal processes. In this study, the supermarket, the laboratory and the hotel buildings are assumed to be installed with DRUPS units, meeting the technical requirements for participation in the FCDM program.

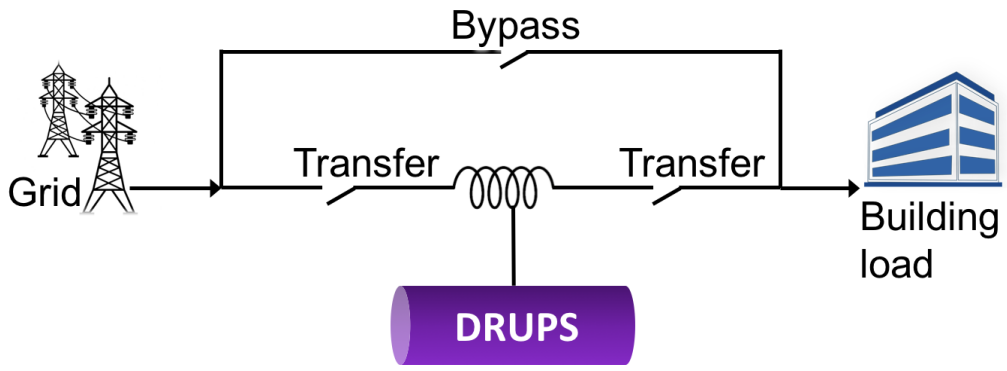

Figure 60: Connection configuration of a DRUPS unit in a commercial building site

5.3.2 Partial load curtailment

A major share of the whole building load in large consumer buildings is constituted by flexible loads such as heating, ventilation and air-conditioning (HVAC) systems. Load share of HVAC systems in a supermarket building is shown Figure 61. Based on field data and previous studies [160–164] (refer Chapter 7 Section 7.2.2 for more details), these loads may provide load curtailment in the range of 10% – 50% of the whole building load. The HVAC loads can be curtailed without affecting the building occupiers' comfort [20]. Some of the HVAC load curtailment strategies adopted in commercial buildings include set-point temperature changes, duty cycling of rooftop units, switching-off supply fans, duct static pressure reduction, chiller ramp down, etc. [162]. Field data show that HVAC loads in such buildings can respond to DR signals well within five minutes [163] and hence meet most of the technical requirements for participation in DR programs [165]. Whole building meter data is easily available from large consumer buildings [166]. It is assumed that, the HVAC loads in the retail store, office and hospital buildings considered in this study, meet the minimum capacity requirement for participation in the ERS program operated by ERCOT.

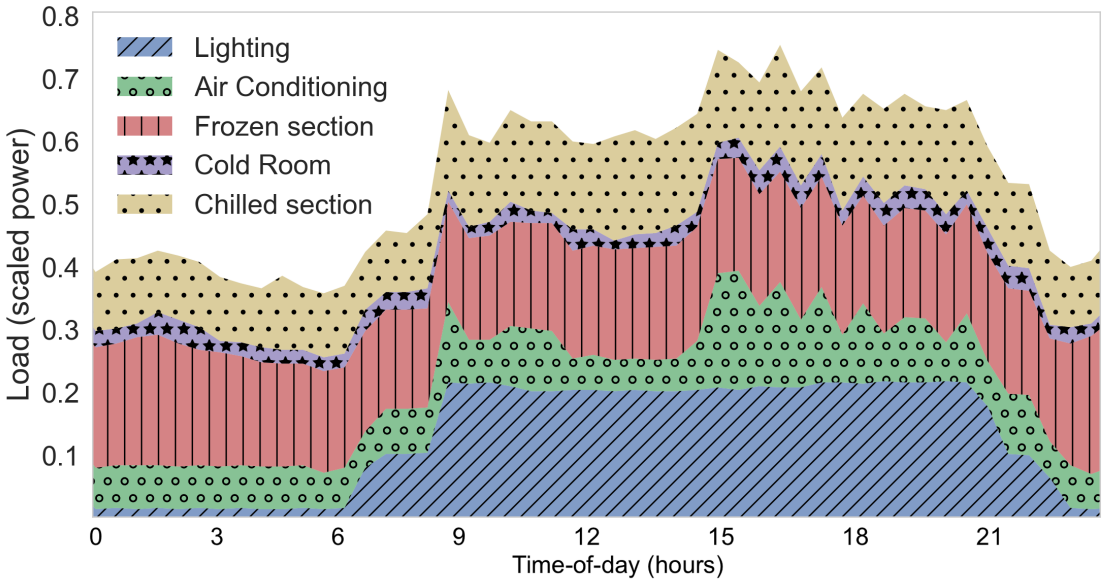

Figure 61: Load share of HVAC systems in a supermarket building

5.4 Data-driven building load estimation

Development of the data-driven capacity scheduling model in FCDM and performance evaluation model in ERS programs follow the modelling approach discussed in Chapter 4. These are discussed below.

5.4.1 Data pre-processing

One year long whole-meter data are collected from the supermarket, the laboratory and the hotel buildings participating in the FCDM program at 30 minutes intervals. Similar timeseries whole-meter data are collected from the retail store, the office and the hospital participating in the ERS program at 15 minutes intervals. No outliers were observed in the meter data and these are scaled using standardisation (discussed in Chapter 4 Section 4.2.5). Single timestamp gaps are filled by imputing with the mean of preceding and succeeding values. Entire data samples (rows) are removed in the case of two or more consecutive gaps.

Load profiling is performed on the whole-meter data obtained. The time-of-day load variations of these building across day-of-week and month-of-year are shown

in Figure 62 and Figure 63. The curves represent annual mean values and the bands represent the 95% bootstrap confidence interval (discussed in Chapter 4 Section 4.2.3.1).

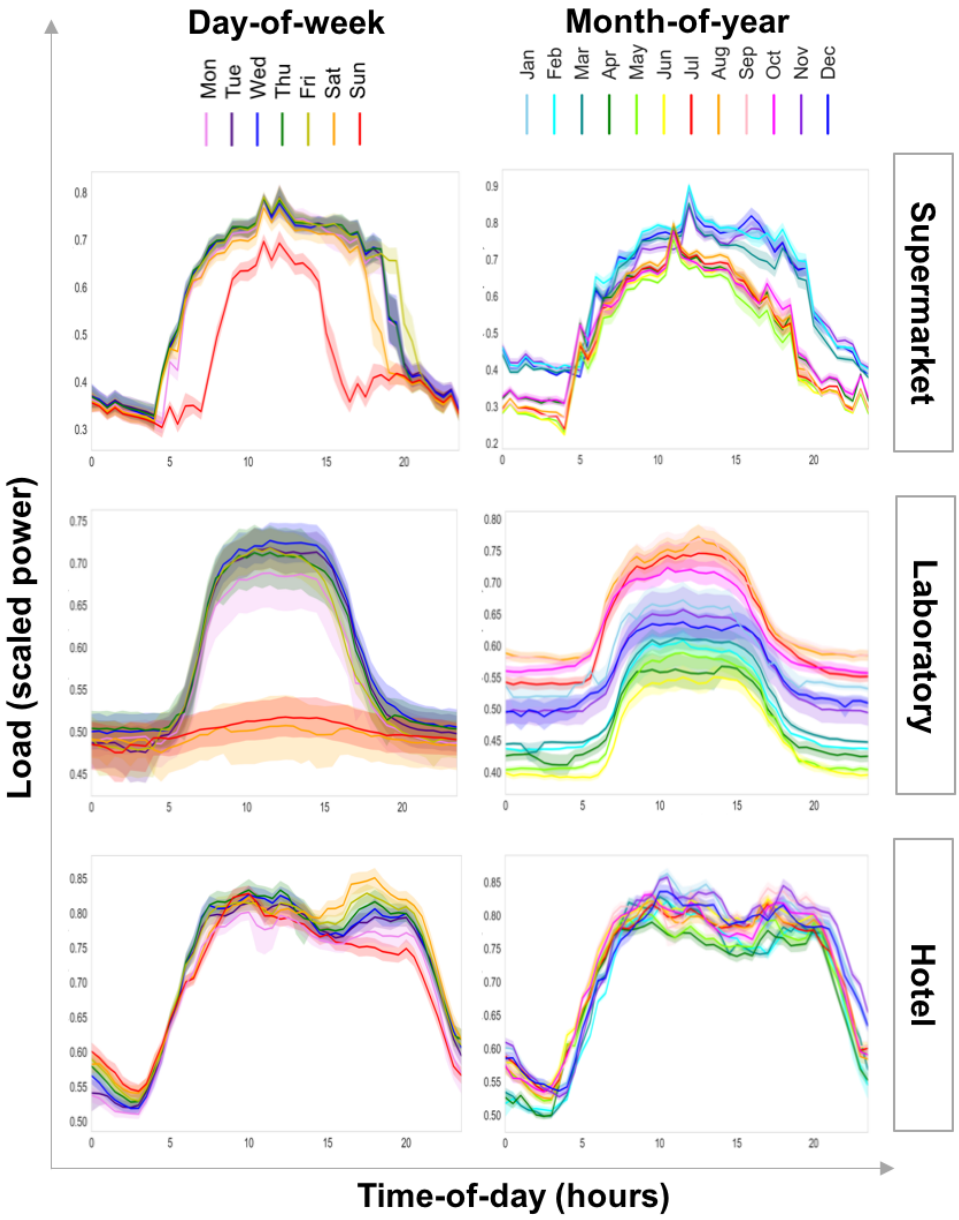

Figure 62: Large consumer buildings in FCDM program (band shows 95% bootstrap confidence interval)

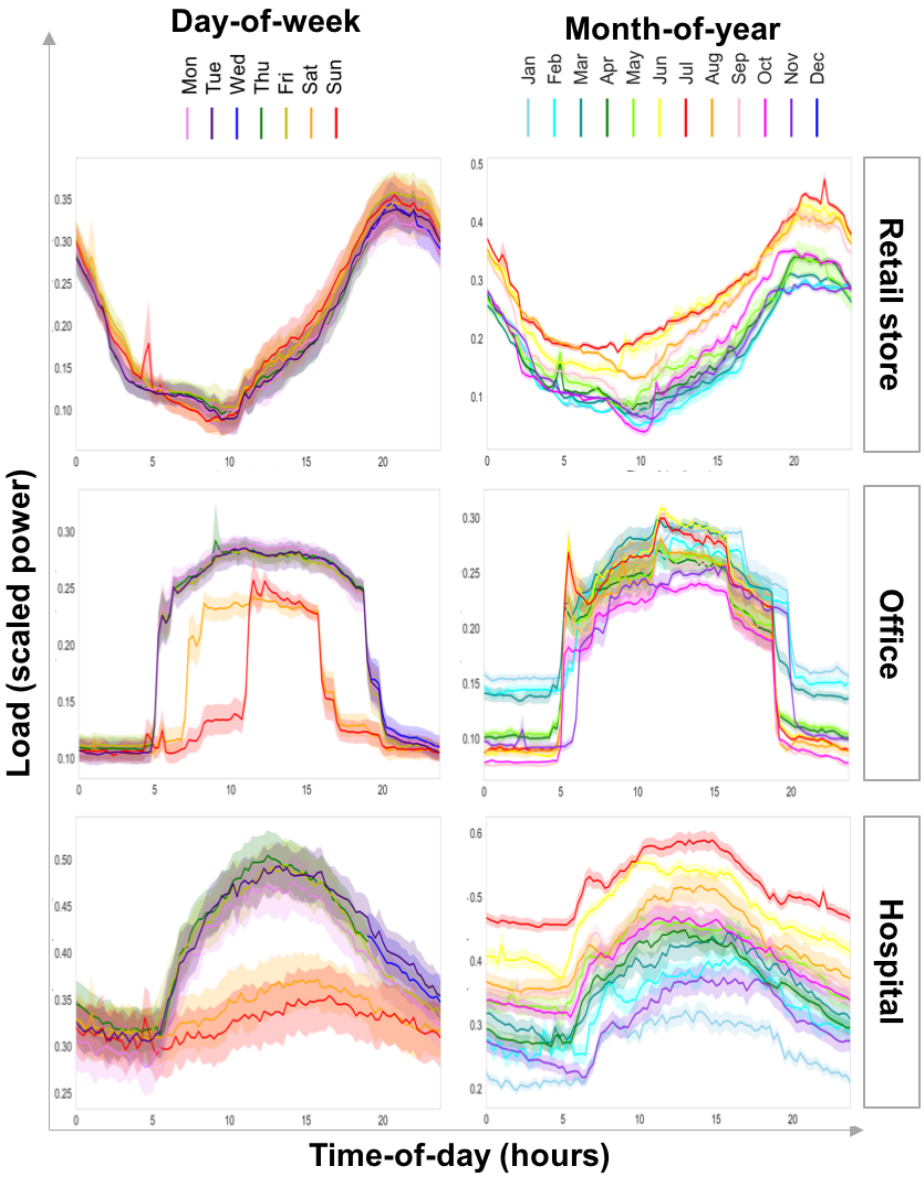

Figure 63: Large consumer buildings in ERS program (band shows 95% bootstrap confidence interval)

The time-of-day variable brings out the diurnal influence of sunshine and temperature on the building load. For example, in the supermarket, the laboratory, the office and the hospital, peak loads occur during daytime and reflects the higher cooling requirement due to increase in outside temperature. The day-of-week variable highlights the impact of occupancy on the building loads. For instance, in the supermarket, Sunday loads are significantly lower than the other days. The month-of-year variable captures the influence of seasonality on the building load.

The three seasons of summer, winter and spring are clearly demarcated for the laboratory building loads, with highest load in the summer.

In this study, only Class I predictors are used. These include temporal predictors such as time-of-day and day-of-week. Datasets for model development are obtained with the temporal predictors and the scaled meter data in timeseries. Weather predictors have not been used to keep the data-driven modelling as close as possible to the conventional modelling approach, discussed later in Section 5.4.3.

5.4.2 Model development

5.4.2.1 Training and testing sets

Based on the size of pre-defined window, the FCDM capacity scheduling model should estimate the whole building load for one week-ahead at 30 minutes resolution and the ERS performance evaluation model for 3 hours at 15 minutes resolution. In ML, this translates to the number of data samples in the testing set (referred to as its size).

Training set samples are taken from the preceding days of the testing set without gaps in between. The training-testing sets are selected for cross-validation in a forward sliding window basis as discussed in Chapter 4 Section 4.3.1.3. The sets slide forward in steps equal to the size of the testing sets. For the FCDM participants, since week-ahead building load estimates (of 30 minutes resolution) are required, the training set sizes are added in weeks. For the ERS participants, since 3 hour building load estimates (of 15 minutes resolution) are required, the training set sizes are added in days. Mean absolute error (MAE) metric is used to measure the training and testing errors of the models.

The size of training sets for each building is chosen with the help of learning curve plots derived using cross-validation of 40 training-testing sets (discussed in Chapter 4 Section 4.3.1.5). The ANN and GBT regression algorithms are used for this purpose. The SVM algorithm is not used because of the higher computational times

required for the model development, as observed in Chapter 4. Based on the learning curves plotted in Figure 64 and the Figure 65, the sizes of the selected training sets for different buildings and algorithms are summarised in Table 11. These range between 3 – 6 weeks for the FCDM participants and between 5 – 6 days for the ERS participants.

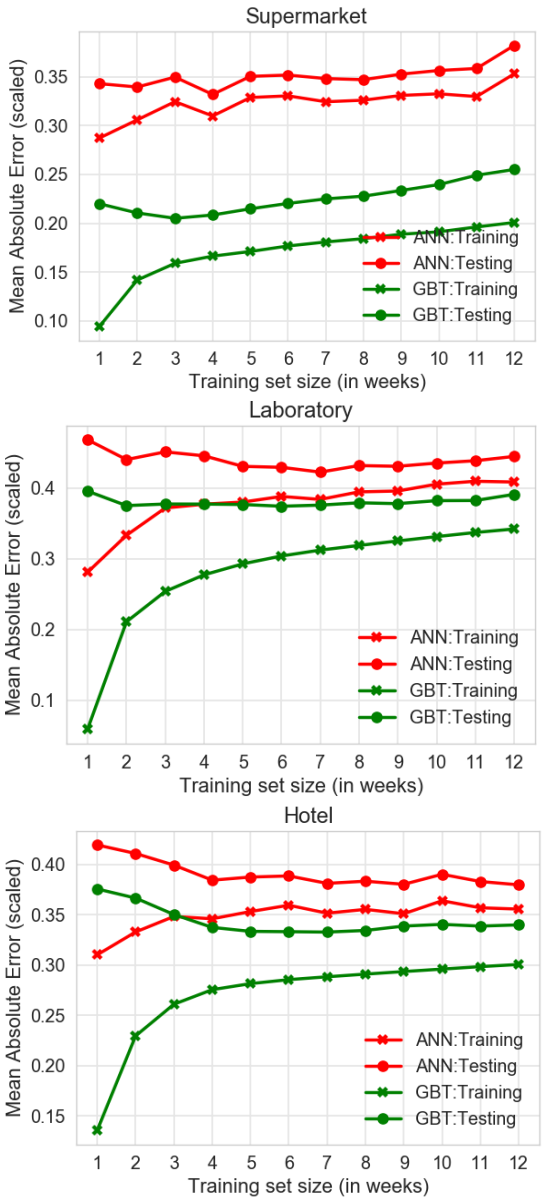

Figure 64: Learning curves for the FCDM participant buildings with testing set size of 1 week and 30 minutes resolution

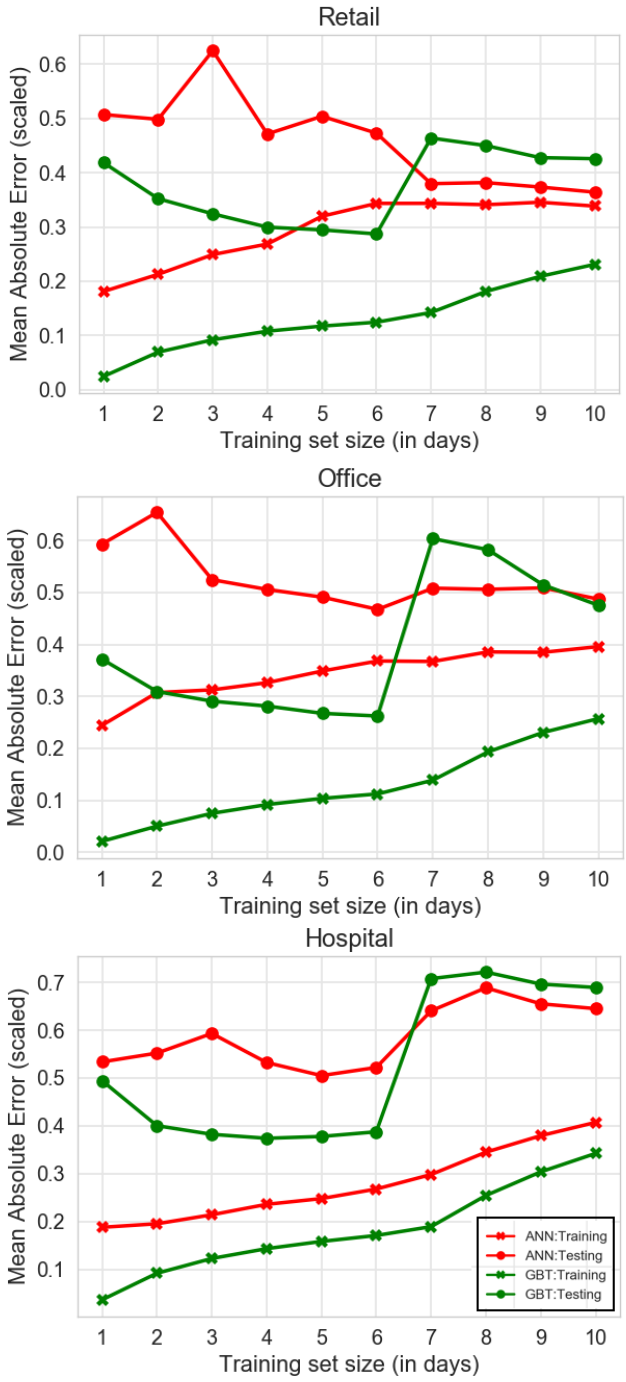

Figure 65: Learning curves for the ERS participant buildings with testing set size of 3 hours and 15 minutes resolution

Table 11: Selected sizes of the training sets for different buildings and algorithms

DR program (data resolution)	Building	Training set size for ANN	Training set size for GBT
FCDM (30 minutes)	Supermarket	4 weeks	3 weeks
	Laboratory	6 weeks	6 weeks
	Hotel	4 weeks	5 weeks
ERS (15 minutes)	Retail	7 days	6 days
	Office	6 days	6 days
	Hospital	5 days	5 days

5.4.2.2 Model calibration, selection and evaluation

The hyperparameters considered for ANN and GBT algorithms are listed in Table 6 (Chapter 4) along with the trial values. Detailed descriptions of these ML algorithms and the hyperparameters were given in Chapter 4 Section 4.3.2. As part of the model calibration, selection and evaluation, the following steps are adopted for each ML algorithm on every building dataset.

- Testing set sizes are fixed based on the DR program specification i.e., 1 week for FCDM and 3 hours for ERS. Training sets are selected from the preceding periods following the sizes given in Table 11. The training-testing sets are selected from the annual dataset on a forward sliding window basis that slides forward in steps equal to the testing set size. From these sets, 40 are sampled and equally distributed into a development set and an evaluation set (20 sets in each). The training-testing sets with UK calendar holidays are avoided from the evaluation set for a fair comparison with the conventional model discussed later in Section 5.4.3. The development set is used for model calibration and selection, which is then validated in the evaluation set.
- Only one round of random-search HPO for model calibration is performed here. During random-search, a sampling size of 75 is used, i.e. 75 random hyperparameter sets are picked from the trial hyperparameter values given in Table 6 to form a search space.

- Cross-validation is performed on the development set using different hyperparameter sets from the search space. The MAE metric is used to measure the testing errors. The hyperparameter set giving the least average MAE on the testing sets is considered to be good. The ML algorithms with the good hyperparameter set are considered as calibrated models on the given dataset.
- The calibrated models are used on the evaluation set for further cross-validation. The models with the least average MAE on evaluation set is selected as the best data-driven model and assigned to the building dataset for the intended DR operation.

The model selection using two ML algorithms (ANN and GBT) for all the six building datasets using the above steps is graphically represented in Figure 66. The figure represents a good set of hyperparameter values for each building dataset within a 2D space of trial hyperparameter values for each ML algorithm. The best model for each building is subsequently selected.

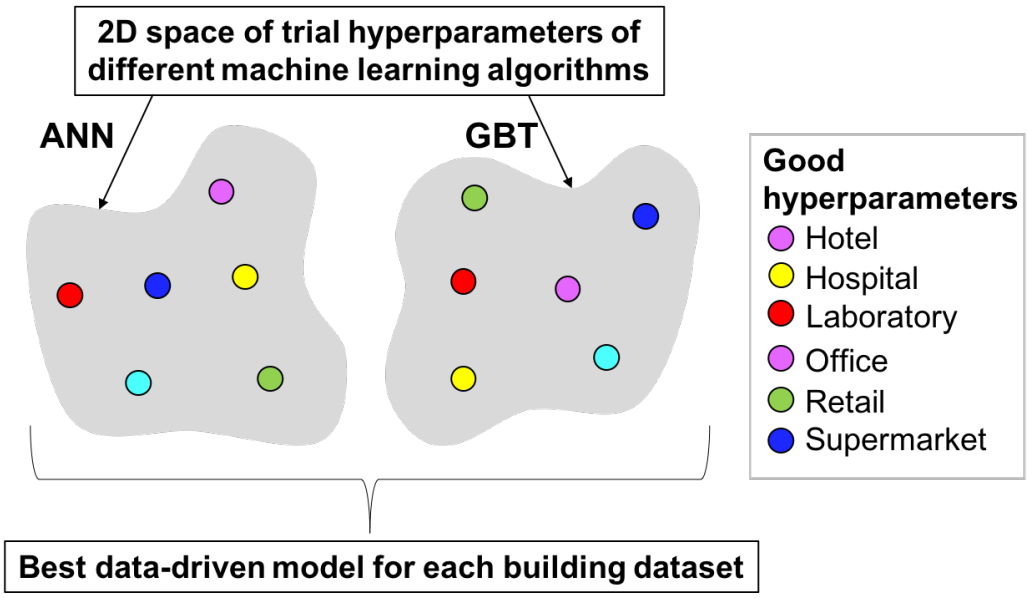

Figure 66: Model calibration for all buildings considered in the study

5.4.3 Comparison with conventional models

The conventional models used for capacity scheduling and performance evaluation in the respective electricity markets are discussed in this section. The methodology of their comparison with the equivalent data-driven models are also elaborated.

5.4.3.1 Preceding week model for FCDM

For DR capacity scheduling, there is no standard model recommended by the FCDM program. However, a common practice among building DR participants is to use the preceding week meter data and commit that for the week ahead. Load variations due to holidays are accounted for, in which case, reference is made from the long term historic data. For instance, the scheduled DR capacity for the upcoming Monday will be the same as the measured load for the previous Monday, unless it is a holiday.

For the purpose of comparison with the data-driven model, meter data for a week are considered as part of a 'testing set' and the preceding week's data are considered as 'forecasts' for the testing set samples. Mean absolute error (MAE) metric is used to calculate the model error. For each FCDM building participant, from their annual meter dataset, 20 testing sets are selected such that the current week and the preceding week do not contain any calendar holidays. The same testing sets were previously used in the evaluation set for the equivalent data-driven models. The calendar holidays are removed because equivalent days' data are not available in the one year long meter dataset. For each FCDM building participant, the average MAE of preceding week models are calculated and compared with that of the data-driven models. In comparison to the data-driven model, since there is no 'learning' occurring in the preceding week model, there is no requirement for a development set.

5.4.3.2 Day averaging model for ERS

For DR performance evaluation, the ERS program in the US has defined many default baseline estimation models, one of them being the *middle 8-of-10 like days.*

This belongs to the category of day averaging models. The rationale behind this method is that the baseline energy consumption (not power) for a DR event day is similar to the energy consumption on the *like day* that occurs closest to the event day. These like days are classified as: 1) weekdays (excluding holidays), and 2) weekends or holidays, belonging to two separate categories. Usually, the like days preceding the event day are used. The 10 initially selected like days exclude those with DR events, notified unavailability periods and any apparent outliers due to metering errors. The daily total energy consumption for each of these 10 days are calculated and days with highest and lowest values are eliminated. The baseline energy consumption per meter data interval (15 minutes) is estimated as the average of energy consumption for the same interval from the remaining 8 eight days.

To improve accuracy, an event-day adjustment scalar is applied. This is calculated as the ratio of actual to baseline energy consumption for an adjustment period of eight 15-minute intervals starting 3 hours prior to the DR event. The event-day adjustment scalar is then multiplied to the unadjusted baseline previously calculated, which may increase or decrease the estimates [167].

Similar to the preceding week model for FCDM, the performance of day averaging models for ERS building participant are compared with the equivalent data-driven models. For this purpose, from the annual meter data of a building, 20 sets of 3 hour periods at 15 minutes resolution are used. These were previously sampled for testing in the evaluation set for the data-driven modelling. The performance of day averaging models on these sets are calculated based on the MAE metric and their average is compared with the average MAE of the best data-driven model, for the respective ERS building participants. Similar to the argument given for the preceding week models in the previous section, there is no requirement of a development set for the day averaging models.

5.5 Results and discussions

Table 12 lists the calibrated data-driven models based on the ANN and GBT algorithms for each building. It can be noticed that the good hyperparameter values differ across the building datasets although the trial hyperparameter values are the same.

Figure 67 compares the average error for ANN, GBT and conventional models on the building datasets, derived using the respective evaluation sets. These results are critically discussed below.

The conventional preceding week model used for FCDM capacity availability in the case of the supermarket, the laboratory and the hotel, shows higher errors in comparison to the data-driven models. The difference in error between conventional and data-driven models is significantly higher for the supermarket building. The preceding week models for participant buildings use previous week's meter data while data-driven models use training data from the previous 3 – 6 weeks. This may have been an advantage for the data-driven models. In order to improve the preceding week model, longer periods may be considered and averaged to get a better estimate. However, it is not certain whether averaging delivers a better estimate and it can only be validated through trial and error. The data-driven modelling presents a computationally efficient and reliable alternative using minimal number of predictors such as time-of-day and day-of-week. As a result, the data-driven models are generally more useful for capacity scheduling in the FCDM program. It is also possible to replicate the data-driven modelling on a large number of FCDM participant buildings. Among the data-driven models, the GBT model performs better than the ANN.

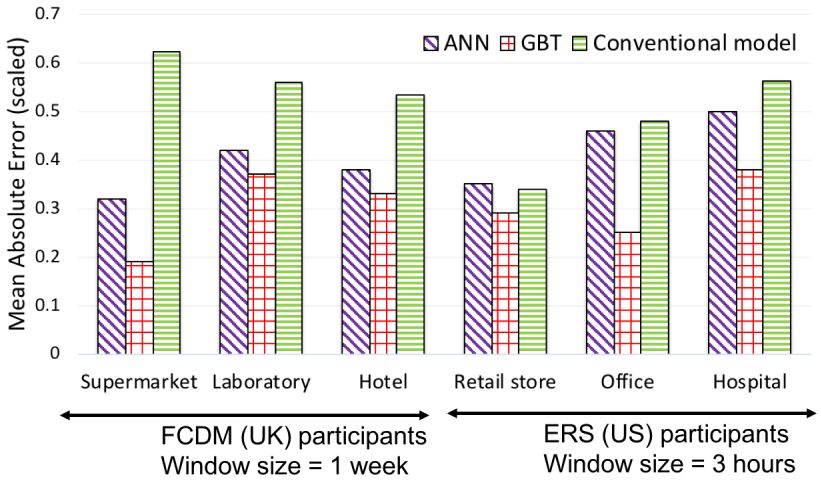

Figure 67: Performance comparison of data-driven and conventional models for the DR participant buildings based on mean absolute error (scaled). A better performing model has lower error

Unlike the preceding week models used in FCDM, the day averaging models in ERS are more competitive with their data-driven alternatives. This may be because of the event day adjustment scalar applied to the day averaging models in ERS. Such measures were missing in the preceding week model used for FCDM. While the day averaging models consider 8 out of 10 like days, the data-driven models use 5 – 7 preceding days for learning the load patterns. The errors between day averaging models and the data-driven models are observed to be generally lower. Moreover, the day averaging model used for the retail store marginally outperforms the ANN model. However, GBT models show superior performance over the ANN and the day averaging models. Different DR program operators in the US consider different variants of the day averaging models such as high 5-of-10 like days (where the 5 highest average energy consumption days are selected), low 4-of-5 like days (where the 4 lowest average energy consumption days are selected) and mid 4-of-6 like days. The selection of the appropriate day averaging model for a given building in a given DR program is more of a heuristic process. In contrast to this conventional approach, the data-driven approach as demonstrated in this study is more reliable for performance evaluation.

Table 12: ML algorithms, hyperparameters, trial and good values for each building datasets

ML algorithm	Hyperparameters	Trial hyperparameter values	Near-optimal hyperparameter values					
			Supermarket	Laboratory	Hotel	Retail store	Office	Hospital
ANN	*Size of hidden layer*	[1][(5,) (5,5), (5,5,5)]	(5,5)	(5,5,5)	(5,5)	(5,5)	(5,5,5)	(4,4,4)
	Activation function	[Logistic, Tanh, Relu]	Tanh	Logistic	Logistic	Tanh	Logistic	Tanh
	Optimization technique	[LBFGS, SGD]	LBFGS	SGD	SGD	LBFGS	LBFGS	SGD
	Regularization	[2]ContDist [0.0001,0.1]	0.07	0.05	0.001	0.02	0.001	0.04
GBT	*Number of base models*	[100,250,500,750]	250	100	750	500	250	100
	Loss function	[LS, LAD, Huber]	LS	Huber	LS	LAD	LS	Huber
	Max. depth of a tree	[3]DiscDist [1,100]	38	41	83	74	98	9
	Min. samples to split	DiscDist [1,100]	56	57	17	62	77	95

[1](4,) is 1 hidden layer with 4 neurons, (5,5) is 2 hidden layers and 5 neurons each; [2]Continuous distribution (range); [3]Discrete distribution (range);

Refer to Chapter 4 Section 4.3.4 for detailed explanations

It could be noted that weather predictors have not been used in the data-driven modelling. This has been done to keep the input data as close to that used by the conventional models, hence promoting a fair comparison. Inclusion of weather predictors such as ambient temperature and its derivatives such as degree days is expected to improve the predictive capability of the data-driven models. The improvement could to be more significant for buildings with higher share of weather dependant loads.

A general observation is that, the GBT consistently outperforms the ANN and the conventional models in terms of the lower errors. The data-driven model calibration and selection process, which is the most computationally intensive stage, takes ~ 25 minutes for the GBT model and ~45 minutes for the ANN model. In terms of predictive performance as well as the computational times, GBT model is preferable over ANN. Once deployed, the building load estimation occurs in less than 8 seconds for either of the data-driven models. More aspects related to model deployment are discussed further.

Capacity scheduling is usually the responsibility of the building consumer whereas performance evaluation is conducted by the program operator. This share of responsibility may vary based on the electricity markets and regulations. Nevertheless, both DR operational tasks equally impact the building consumer and the program operator. Regardless of who performs the deployment of the data-driven models, the deployment guidelines remain the same.

General aspects related to ML model deployment were discussed in Chapter 4 Section 4.4. The ML model development performed was particularly focussed on its final deployment, such as the selection of testing set sizes following the DR program windows. Also, the selection of supervised ML algorithms is motivated by their applicability to building load data as well as their minimal usage of computational resources for easy replicability. The data pre-processing steps discussed as part of the modelling, such as data collection and data cleaning will be part of the data flow in the deployed model. Further, it is also important to track the predictive

performance of the model and compare that with the benchmark set during the ML model development.

Figure 68 shows the load variations for other large consumer buildings such as a warehouse, a datacentre and an airport. The data-driven modelling approach developed in Chapter 4 and demonstrated in the presented chapter for application in DR operational tasks such as capacity scheduling and performance evaluation, can be extended to such buildings. This is considered as future work.

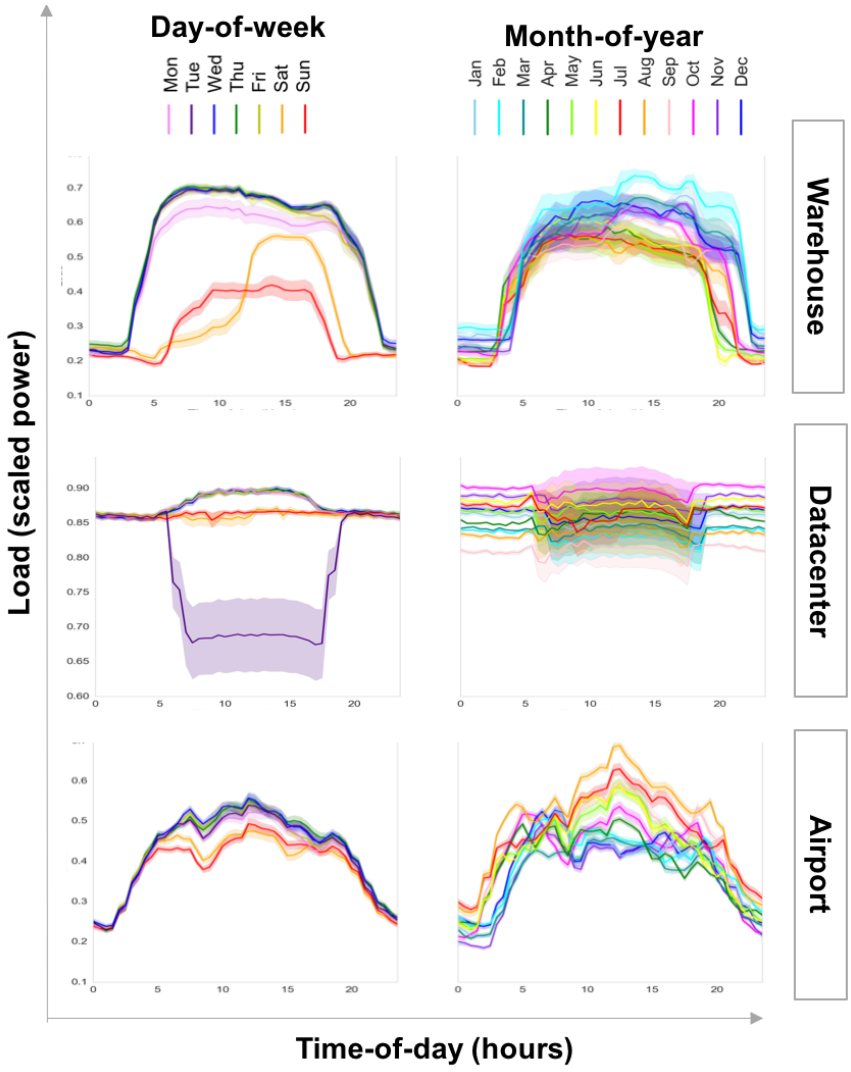

Figure 68: Large consumer building load variations (bands shows 95% bootstrap confidence interval)

5.6 Conclusions

The conventional models widely used for building load estimation in DR operational tasks such as capacity scheduling and performance evaluation in many electricity markets were found to be less reliable. From the system operator perspective, errors in the building load estimates for consumers participating in DR programs would creep into the system planning process. In aggregated scales, this affects the overall grid reliability. From the consumer perspective, the building load estimation errors (if under-estimated) would result in revenue losses. For example, an under-estimated baseline load when compared with the curtailed load fetches a lower utilisation payment than expected. The dependence on unreliable conventional models would discourage the participation of building consumers in DR programs.

The study showed the potential of data-driven models for deployment in DR operational tasks in the context of large consumer buildings participating in ongoing DR programs. The week-ahead load estimation models developed for DR capacity scheduling of a supermarket, laboratory and hotel buildings outperformed the conventional model that is based on previous week's meter data. These buildings were assumed to have full load curtailment capability based on standby generators and participate in the FCDM program in the UK electricity market. The 3 hours ahead baseline load estimation models developed for performance evaluation of the hotel and hospital buildings outperformed the conventional day-averaging model. The hotel, retail store and hospital buildings were assumed to have partial load curtailment capability with flexible loads and participate in the ERS program in the US electricity market.

The demonstrated approach enables development and deployment of customised load estimation models on many large consumer buildings participating in DR programs. This approach may also be extended to industrial buildings with predictable load patterns. Although specific cases of applications such as DR

capacity scheduling and performance evaluation were demonstrated, the data-driven modelling could be used for other DR operational tasks. The conventional models used in DR operational tasks such as capacity scheduling and performance evaluation were developed when the number of participants in DR programs were small. With the increasing penetration of renewables and electric vehicles, the need for DR capacity is escalating and more building consumers are participating in DR programs. Data-driven modelling presented in this chapter enables reliable, fast and cost-effective development of customised models for these large number of participants. The pipeline of data-collection, model development and model deployment could pave the way for automation in the DR industry.

Chapter 6

Data-driven SoC management of a large consumer BESS for reliable operation in multiple DR programs

6.1 Introduction

In this chapter, the supervised machine learning (ML) based building load estimation modelling developed in Chapter 4 is applied in the context of state-of-charge (SoC) management for battery energy storage system (BESS) based demand response (DR). This has resulted in the development of a novel data-driven SoC management strategy that is beneficial to the BESS owner as well as the grid. The developed strategy is simulated and evaluated on a large consumer BESS participating in multiple DR programs.

6.1.1 Background

Energy storage is a necessary resource for the modern electricity grid. A BESS can provide bidirectional power flow with high efficiency, sub-second response time, fast ramp rate and long lifetime [168]. Unlike the pumped water storage systems, there is no locational restriction to BESS in the grid. Battery prices, particularly that of the Lithium-ion (Li) chemistry have been dwindling in the global market [31]. Based on these factors, system operators encourage BESS participation in services such as frequency response, energy arbitrage, peak shaving and congestion management through demand response (DR) programs. BESSs are expected to form an integral part of the smart grid [169].

Many standalone BESSs have already been commercially deployed for grid services in the last decade [168,170,171]. Although not a substantial storage asset at this point of time, the escalating sales of electric vehicles (EVs) and development in concepts such as vehicle-to-grid (V2G) promises large aggregated energy storage capability in the grid [172,173]. BESSs collocated with renewable generation such as solar and wind are on the rise. BESS on the consumer-side (also referred to as behind-the-meter) that are capable of providing system demand peak shaving and electricity cost reduction through load curtailment have started gaining popularity recently [174]. A 2017 survey in the UK reported that investors prioritised consumer-side BESS (49%) over standalone (24%) and collocated (27%) projects [107].

One of the major concerns with regards to reliable operation of a BESS in a DR program is the energy capacity limit. Since ancillary services such as frequency response are not zero-mean (unequal high and low frequency deviations) and batteries are not 100% efficient, energy capacity limits are easily reached [110,111]. Using larger energy capacity results in higher capital costs. Hence it is important to develop strategies to utilise the available energy capacity in favour of the BESS owner as well as the system operator, within the BESS's technical constraints and allowances of the grid.

The SoC is the percentage of the energy remaining in a battery with finite energy capacity. For reliable operation of a BESS in a DR program, an SoC management strategy needs to be adopted. The strategy may differ based on the type of DR program as well as the type of BESS (standalone, EV, collocated or consumer-side). Although not explicitly referred to as SoC management, studies have explored this in the literature. For better comprehension, these strategies are categorised as follows.

- ***Deadband strategy:*** The BESS is recharged using a bias power when frequency is within a non-critical frequency window or deadband around the nominal frequency. This strategy was explored in [108] and [109] for BESS participation

in frequency response programs. The SoC was shown to sustain for longer periods, with lower battery degradation, longer lifetime and higher revenue.

- ***Real-time offset strategy:*** Here, the SoC is maintained based on the offset power purchased through balancing market, trading and other services that do not interfere with the delivery of the contracted service (DR program). The variants of real-time offset strategies were explored in [110–115] in different electricity markets. In general, these strategies are shown to benefit the BESS owner without affecting grid reliability.
- ***Scheduling strategy:*** This considers BESS charging at regular time intervals according to the service it delivers. In [116,117], the scheduling strategy was studied in the context of peak shaving services for buildings where the SoC is adjusted prior to providing backup to the building load and later.
- ***EV strategies:*** SoC management in the context of EV based grid balancing have been explored in [118–120]. Compared to standalone and consumer-side BESS, the EV batteries have smaller energy capacities. Hence the EV SoC management strategies include added constraints of availability and aggregation.

A detailed review of the above literature is provided in Chapter 3 Section 3.4.2. Real-time offset based SoC management interacts continuously with the energy balance of the grid and adds to the uncertainties in planning [113]. Although beneficial if implemented wisely, such strategies have not yet been adopted for SoC management in BESS based DR programs. Since buying price of electricity is always higher than selling price, SoC management strategy based on scheduling is not necessarily in favour of the BESS owner. Further, network capacity constraints may limit the purchase of energy prior to a system demand peak [117]. Deadband SoC management has minimal impact on the grid while enabling the BESS to deliver the services reliably. For the same reason, this strategy has the highest possibility of being adopted for any BESS based DR programs.

6.1.2 Contribution

State-of-the-art SoC management strategies such as the deadband strategy, the real-time offset strategy and the scheduling strategy are rule-based. In practical DR operational scenarios, rule-based SoC management strategies may not always be beneficial for the BESS owner nor for the grid. This is best explained with the following example.

In the case of a large consumer building BESS participating in multiple DR programs such as frequency response as well as evening peak shaving (by providing backup to the building load), a combination of existing deadband and scheduling SoC management strategies may be adopted. Here, the deadband strategy enables reliable delivery of frequency response, which is then interrupted at a scheduled time prior to the daily evening peak, in order to increase the SoC to a level that is sufficient to backup the building load during the peak period. This limits the frequency response participation of the BESS during the scheduled recharging period prior to the evening peak, resulting in revenue loss for the consumer. If multiple units of consumer-side BESS simultaneously perform the scheduled recharging, they would burden the grid.

Data-driven thinking relies on the information derived from data. The decisions made through data-driven thinking adapts to new data, rather than conforming to pre-set rules. In the scenario discussed above, it is possible to design an SoC management strategy that adapts to the building load for the daily evening peak period. ML models trained online using incoming data are useful for adaptive building load estimation. A modelling approach based on supervised ML was discussed in Chapter 4. Daily recurring building load forecasts for the peak shaving period enable the implementation of a data-driven SoC management strategy that is beneficial for the BESS owner as well as the grid. Such a data-driven SoC management strategy is developed and demonstrated in this chapter.

In this study, two ongoing DR programs in the UK market are considered, 1) enhanced frequency response (EFR) [23] – where the BESS is required to provide symmetric, dynamic and fast frequency response to the grid and 2) red zone management (RZM) – where BESS building load backup is provided to avoid the evening system demand peak price (red zone). The data-driven SoC management strategy is developed based on the framework of the EFR and RZM programs. The building load forecast model used in the data-driven SoC management strategy is developed based on a gradient boosted trees (GBT) ML algorithm [30,31]. The economic viability of a university building BESS participating in the DR programs operating based on the developed strategy is assessed through a case study. A sensitivity analysis has also been performed to test the robustness of the strategy.

To the best of our knowledge, a data-driven SoC management strategy has not yet been studied in the literature. In the presented study, such a strategy has been developed to ensure reliable operation of a large consumer building BESS in two DR programs without impacting the consumer nor the grid. Although the DR programs considered in this study and their requirements may change in the future [32], the developed strategy could motivate further research in data-driven modelling for BESS based DR.

6.1.3 Layout

The rest of the chapter is structured as follows. Section 6.2 discusses the existing service framework for EFR and RZM DR programs. Section 6.3 develops the data-driven SoC management strategy. Section 6.4 details the methods used for economic analysis of the consumer building BESS in multiple DR programs. Section 6.5 includes a case study on a university building BESS participating in the DR programs. The developed data-driven SoC management strategy is simulated in this case study and a sensitivity analysis is also performed to test the robustness of the results. This is followed by an economic analysis. Section 6.6 is a critical discussion on the assumptions and results. Section 6.7 concludes the chapter.

6.2 DR program framework

6.2.1 Enhanced frequency response (EFR)

The UK electricity grid has a nominal frequency of 50 Hz and a statutory frequency deviation limit of $\pm 500\, mHz$. A first of its kind DR program, the EFR was introduced in 2016 to enable BESS participation in frequency responsive DR programs. This section discusses the EFR program framework based on the guidelines given by National Grid (NG), the UK transmission system operator (TSO), here [33].

According to the EFR program framework, a narrow deadband of $\pm 15\, mHz$ and a wide deadband of $\pm 50\, mHz$ are allowed for deadband SoC management. The participating BESS is required to provide dynamic high and low frequency response, outside the deadband but within the statutory deviation limits ($\pm 500\, mHz$). No response (zero operational power) is expected from the BESS when the frequency is within the deadband. The dynamic aspect of the EFR program represents the linear increase of the response from the deadband limits to the statutory limits (contracted maximum operational power) for high and low frequency deviations. When frequency is on the higher side, energy is absorbed into the BESS and vice-versa. The frequency response is expected to reflect on the grid within 1 second, which should include the frequency signal detection, communication and reaction time of the BESS.

Although a more demanding SoC management allowance for the BESS, the narrow deadband is considered in this study due to its higher value to the BESS owner as well as the grid. Figure 69 shows the EFR service profile for narrow deadband on a power-frequency plot. The purple line represents a *service reference* and the black lines represent a *service envelope*. Based on a maximum operational power of 100% (+ve if export and –ve if import), the service profile is discussed further.

The service reference serves the purpose of providing a target for participants who have infinite energy capacity and do not need to manage the SoC. In this case, the operational power and ramp rate (power/second) is 0% within the deadband of $\pm 15\ mHz$. Outside the deadband, the operational power should reach $\pm 100\%$ at the statutory frequency deviation limit of $\pm 500\ mHz$, at an operational ramp rate of:

$$\left(\frac{-1}{(0.500-0.015)} {}^{df}/_{dt}\right).100\%$$

$$\text{or equivalently } \left(\frac{-1}{0.485} {}^{df}/_{dt}\right).100\% \tag{6.1}$$

where df is the frequency deviation (Hz) in the time interval dt (seconds).

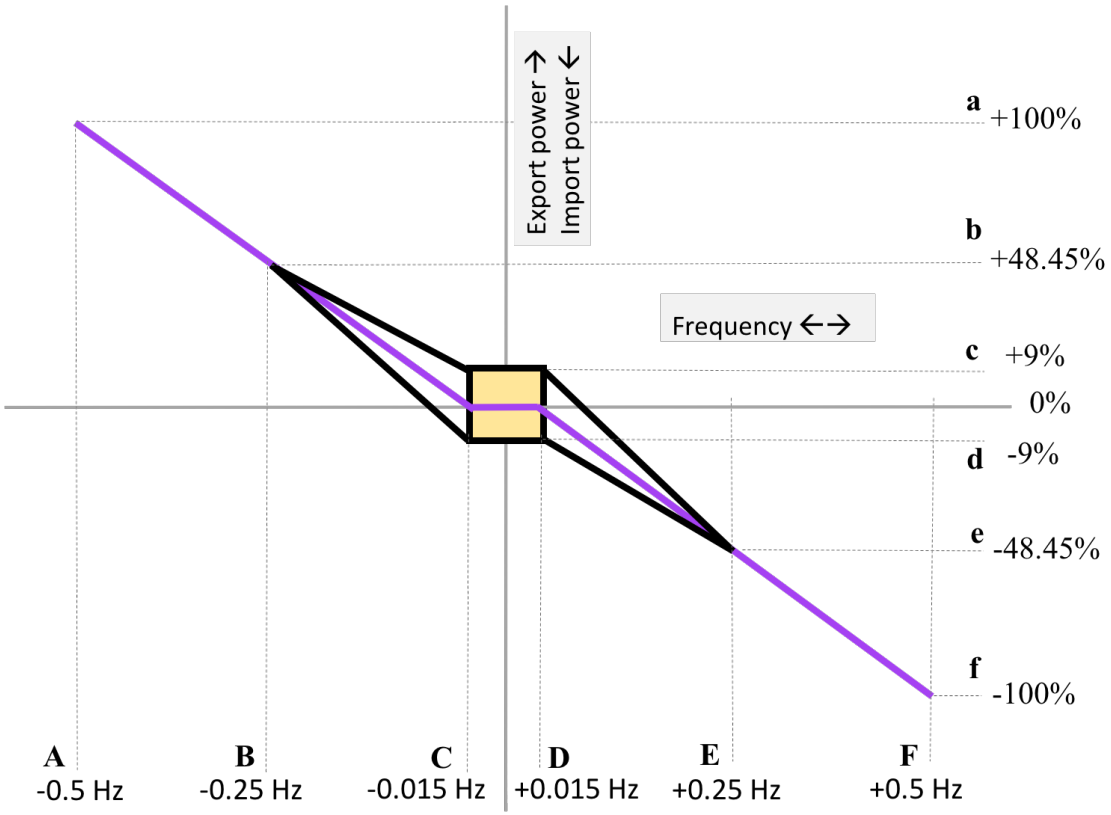

Figure 69: Enhanced frequency response service delivery profile (recreated from [33])

The service envelope allows more flexibility while delivering EFR. The envelope includes a deadband zone (square shaded area between C and D) within which no frequency response is expected and the BESS may vary the operational power

between $\pm 9\%$ using a slow operational ramp rate of $\pm 1\%$. The area of the envelope excluding the deadband zone, upto the frequency deviation of $\pm 250\ mHz$, allows a ramp rate of:

$$\left(\frac{-1}{0.485} {}^{df}/_{df \pm 0.01}\right).100\% \tag{6.2}$$

The service envelope provision enables SoC management for finite energy capacity systems such as BESS. Outside the frequency deviation of $\pm 250\ mHz$, the service reference ramp rate is to be followed until the statutory limits of $\pm 500\ mHz$ are reached. The BESS is expected to have an energy capacity that can deliver the maximum operational power at $\pm 500\ mHz$ for 15 minutes.

6.2.2 Red zone management (RZM)

In the UK, the electricity distribution infrastructure is operated and owned by the distribution system operators (DSOs). The cost of operating and maintaining a reliable infrastructure is recovered from the consumers through the distribution use of system (DUoS) tariffs, based on time-of-use (ToU). Since the characteristics of the network and consumers are different across the DSOs, these tariffs also vary. Figure 70 shows the ToU based DUoS tariff for large consumers in the South-East England [175].

The green zone (23:00 – 7:00) has the lowest tariff, followed by the amber zone (07:00 – 16:00 & 19:00 – 23:00). The highest tariff appears during the red zone (16:00 – 19:00) which coincides with the system peak demand period. The remarkably high tariff in the red zone encourages consumers to reduce or shift consumption, in turn helping shave off the system peak. The RZM program in the UK aims at peak shaving through ToU tariffs for large consumers such as commercial buildings. This could be categorised as a price-based DR program. The DUoS tariffs in the South East England are used as a reference case.

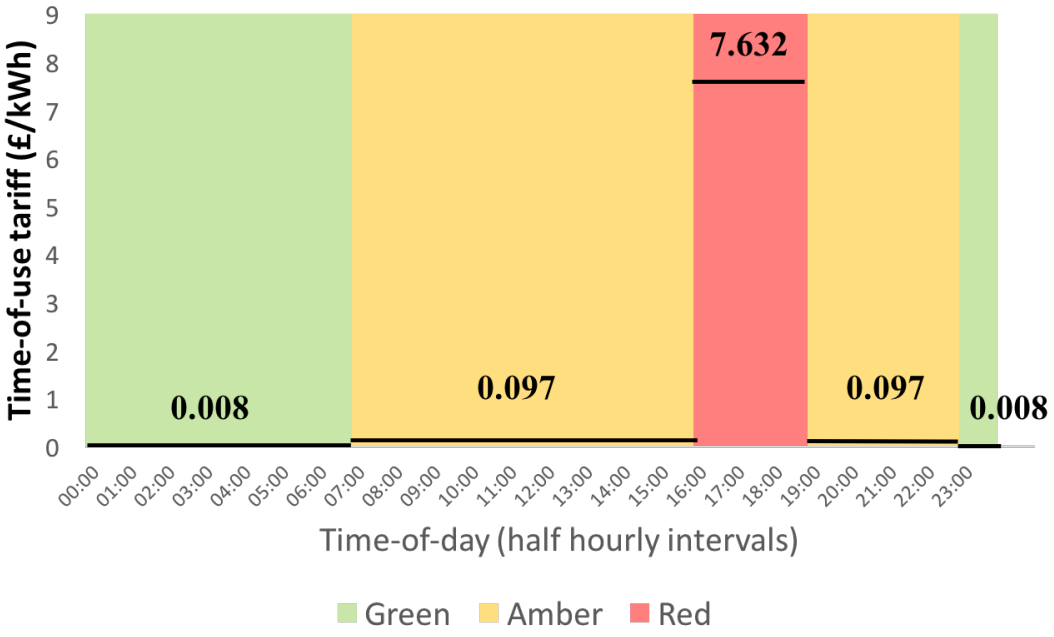

Figure 70: DUoS tariffs based on ToU for large consumers in the South-East England

6.3 SoC management strategy development

6.3.1 Deadband SoC management in EFR

As discussed in Section 6.2.1, during deadband SoC management for EFR, the operational power is allowed to vary between ±9% using a slow operational ramp rate of ±1%. The SoC targeted by the controller in the BESS, as part of the SoC management strategy is referred to as the set-point SoC, SoC_{sp}. When frequency is within the deadband, the operational power as seen on the grid at time s is:

$$P_{grid,s} = \begin{cases} P_{grid,s-ds} - 1\%.P_{gridmax} & ; 0 \leq \mathrm{SoC_s} \leq \mathrm{SoC_{sp}} \; ; \mathrm{P_{grid,s}} > -9\%.\mathrm{P_{gridmax}} \\ P_{grid,s-ds} + 1\%.P_{gridmax} & ; \mathrm{SoC_{sp}} < \mathrm{SoC_s} \leq 100 \; ; \mathrm{P_{grid,s}} < 9\%.\mathrm{P_{gridmax}} \\ -9\%.P_{gridmax} & ; \mathrm{P_{grid,s}} \leq -9\%.\mathrm{P_{gridmax}} \\ 9\%.P_{gridmax} & ; \mathrm{P_{grid,s}} \geq 9\%\mathrm{P_{gridmax}} \end{cases} \quad (6.3)$$

where, ds is the time resolution of frequency data, $P_{grid,s-ds}$ is the operational power at time $s-ds$, $P_{gridmax}$ is the maximum operational power, SoC_s is the SoC at time s. The SoC of the BESS at time s is updated as,

$$SoC_s = SoC_{s-ds} - \frac{1}{E_{battery}}.\begin{cases} \left(\frac{P_{grd,s}}{\eta_d}\right).ds & ; \text{ export to grid} \\ \left(-P_{grid,s}.\eta_c\right).ds & ; \text{ import from grid} \end{cases} \quad (6.4)$$

where $E_{battery}$ is the BESS energy capacity, η_c is the charging efficiency, η_d is the discharging efficiency. For a set-point SoC of 50%, the EFR deadband SoC management allowance is visualised in Figure 71.

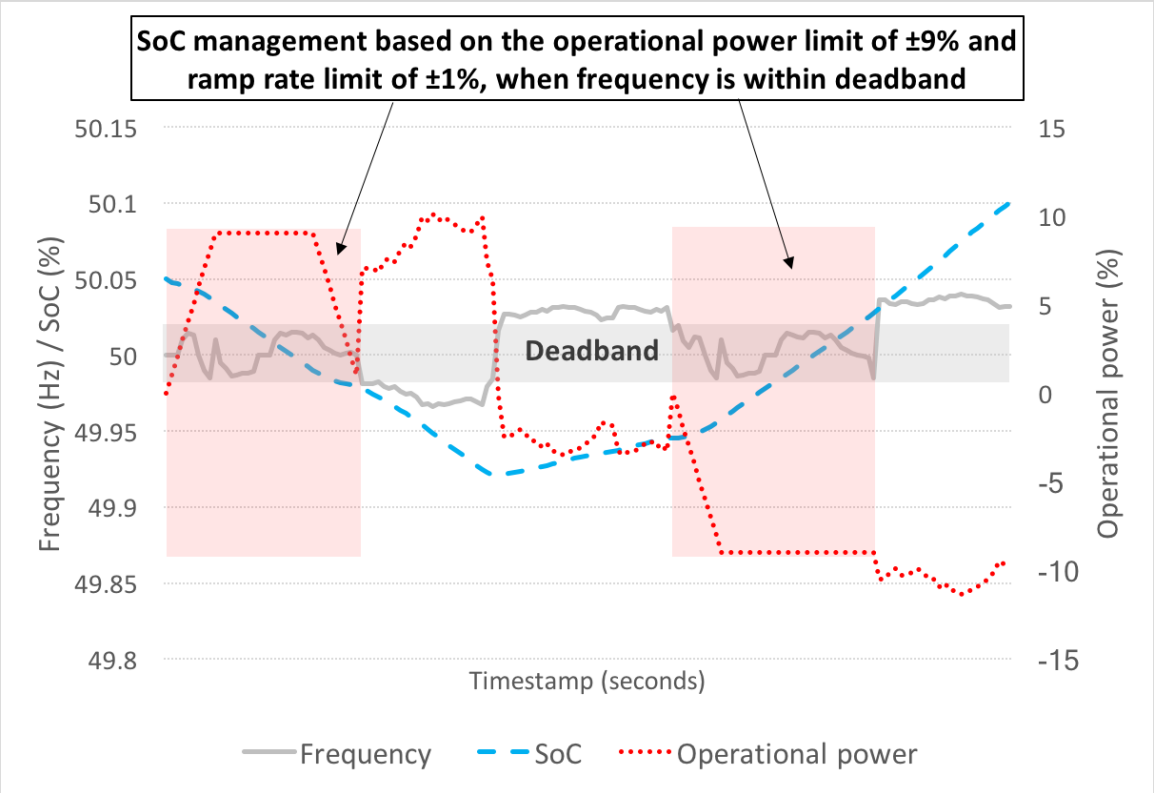

Figure 71: Deadband SoC management in the EFR program with set-point SoC of 50%

While participating in the EFR program exclusively, it is ideal to maintain the BESS set-point SoC at 50%, so that there is sufficient energy capacity allowance for high and low frequency response. Based on the actual recorded frequency data in the UK

electricity system (2015) and the service profile of the EFR program, frequency response (operational power) is demonstrated in Figure 72. This assumes an energy/power ratio of 5.

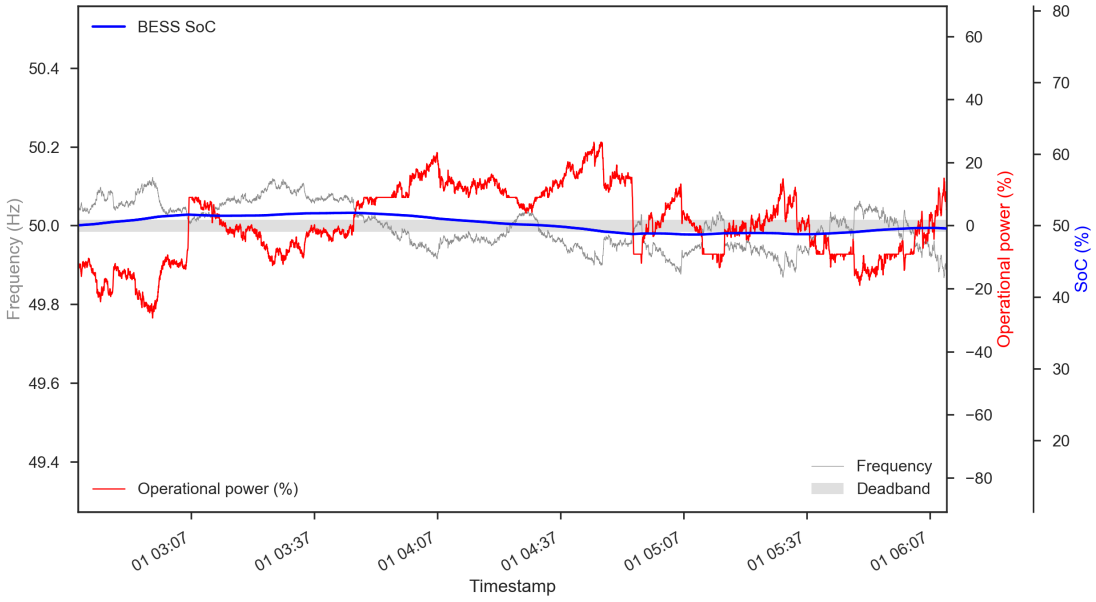

Figure 72: BESS based dynamic frequency response based on the EFR program framework (SoC wiggles around the set-point value of 50%)

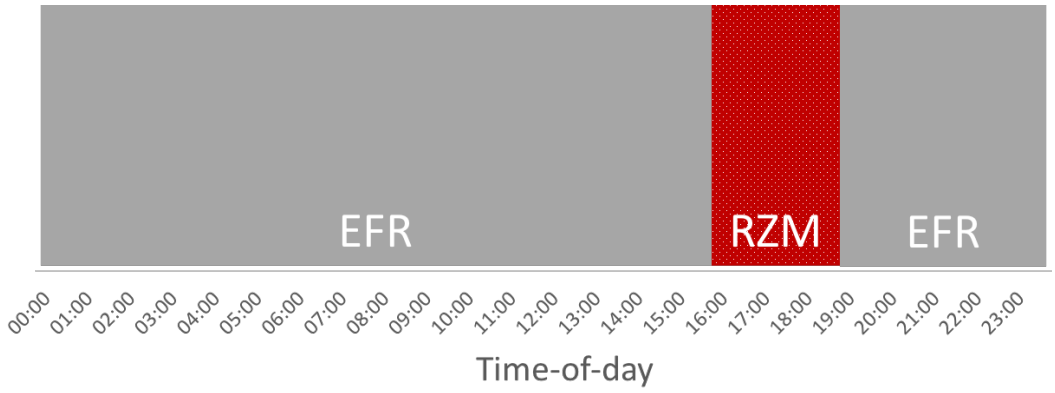

Figure 73: Participation windows considered for EFR and RZM DR programs

6.3.2 SoC management for EFR and RZM participation

Participation windows considered for the EFR and RZM DR programs are shown in Figure 73. The RZM delivery window is 16:00 - 19:00 hours and EFR is delivered during rest of the hours. The BESS rated power should be larger than the building peak load during these hours. Similarly, the BESS energy capacity should be larger than the maximum energy required to provide a 3 hour backup to the building load during the RZM period.

It was mentioned that 50% is the recommended set-point SoC for a symmetric frequency response service such as EFR. Starting at 50% SoC prior to the RZM period (pre-RZM) would result in near-depletion of the SoC by the end of this period, since the BESS energy is provided as building load backup. In this case, the BESS needs to be recharged back to 50% SoC after the RZM period (post-RZM) to continue the EFR participation.

Figure 74 shows the frequency duration curve based on actual recorded frequency datasets from the UK grid. For either of the datasets, majority of the frequency deviations are within $\pm 100\ mHz$ and large frequency deviation events are sparse. Further, $-100\ mHz$ deviations (low frequency response) are of longer duration compared to $+100\ mHz$. Hence, the deviations within $\pm 100\ mHz$ are not zero-mean and shows a consistent bias towards low frequencies. It could be concluded that, for symmetric (high and low) frequency response services such as EFR, it is possible to take more risk with higher SoC (above 50%) than with lower SoC (below 50%), so as to avoid service delivery failures. Hence, it is recommended to achieve a higher set-point SoC in the pre-RZM period, rather than doing the recharge in the post-RZM period.

A higher set-point SoC could be achieved in the pre-RZM period, using scheduled recharging at the maximum possible ramp rate. However, this faces three issues: 1) EFR participation is not possible during this period resulting in revenue loss for the

BESS owner, 2) costs of buying electricity are higher and 3) multiple BESS units performing the scheduled recharging may burden the grid.

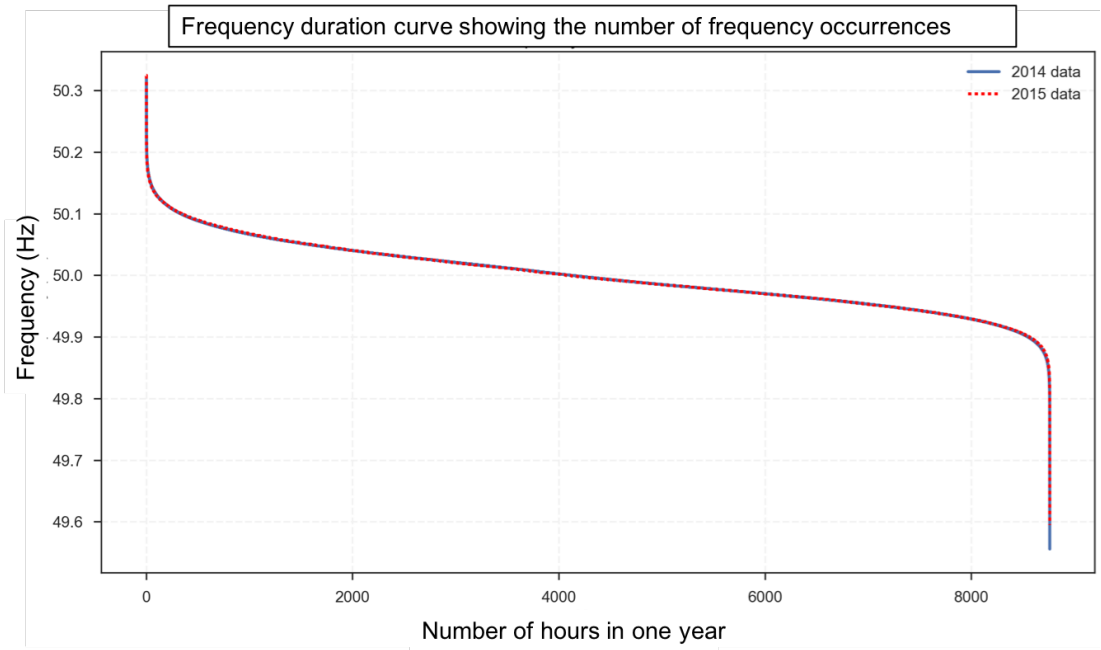

Figure 74: Frequency duration curve using actual recorded data in the UK electricity grid

So as to avoid these issues, a data-driven SoC management strategy is adopted using the narrow deadband allowance in the EFR program within the pre-RZM period. This is achieved through estimation of a data-driven set-point SoC and pre-RZM recharge period discussed further.

6.3.2.1 Data-driven set-point SoC estimation

Building energy backup for the RZM period (3 hours) is estimated one day-ahead using a data-driven load estimation model. The model is developed using the approach discussed in Chapter 4. Based on the required energy backup (E_{backup}) estimated using the data-driven model, a data-driven set-point (ddsp) SoC is derived using the following equation:

$$SoC_{ddsp} = SoC_{sp} + 0.5 * \frac{E_{backup}}{E_{battery}} * 100\% \quad (6.5)$$

where SoC_{sp} is the set-point SoC in EFR only scenario (i.e. 50%) and $E_{battery}$ is the BESS energy capacity. In the equation, the 0.5 factor is used to equally distribute the backup energy requirement around SoC_{sp}.

The data-driven set-point SoC, SoC_{ddsp} is used as the new set-point for SoC management within the EFR window, in a pre-RZM recharge period. The BESS continues to deliver EFR during the recharge period while aiming to achieve SoC_{ddsp}. This is represented in Figure 75. Since occurrences of high frequency deviations are lower, it is possible to take the risk of EFR service delivery failure during this period.

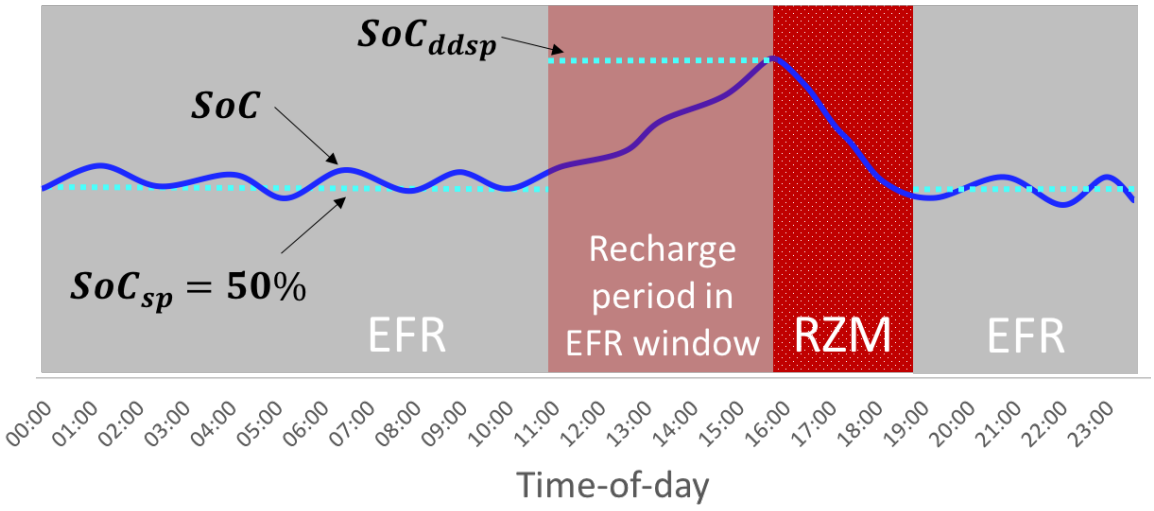

Figure 75: BESS EFR and RZM participation based on the data-driven set-point SoC

6.3.2.2 Recharge period estimation

The duration of the recharge period depends on the SoC_{ddsp}, the response to the frequency signals that occur during this period as well as the energy/power ratio of the BESS.

Accurate estimation of SoC_{ddsp} enables better day-ahead planning for the pre-RZM recharge period. Also, these estimates need to be made on a continuous operational set-up. Data-driven models for building load estimation are useful in this regard.

Figure 74 shows that most of the frequency deviations in the UK grid are within $\pm 100\ mHz$ over a year. There is consistency in these small frequency deviations even across two different years. Further, large frequency deviations are sparse. As a result, most of the times, the BESS SoC shows a near-linear trend while trying to achieve the SoC_{ddsp} within the pre-RZM period.

Based on actual frequency signals recorded (2015) in the UK electricity grid, the recharge periods for BESSs with different energy/power ratios are shown in Figure 76. The figure shows different BESSs trying to achieve $SoC_{ddsp} = 100\%$, starting from the default 50%.

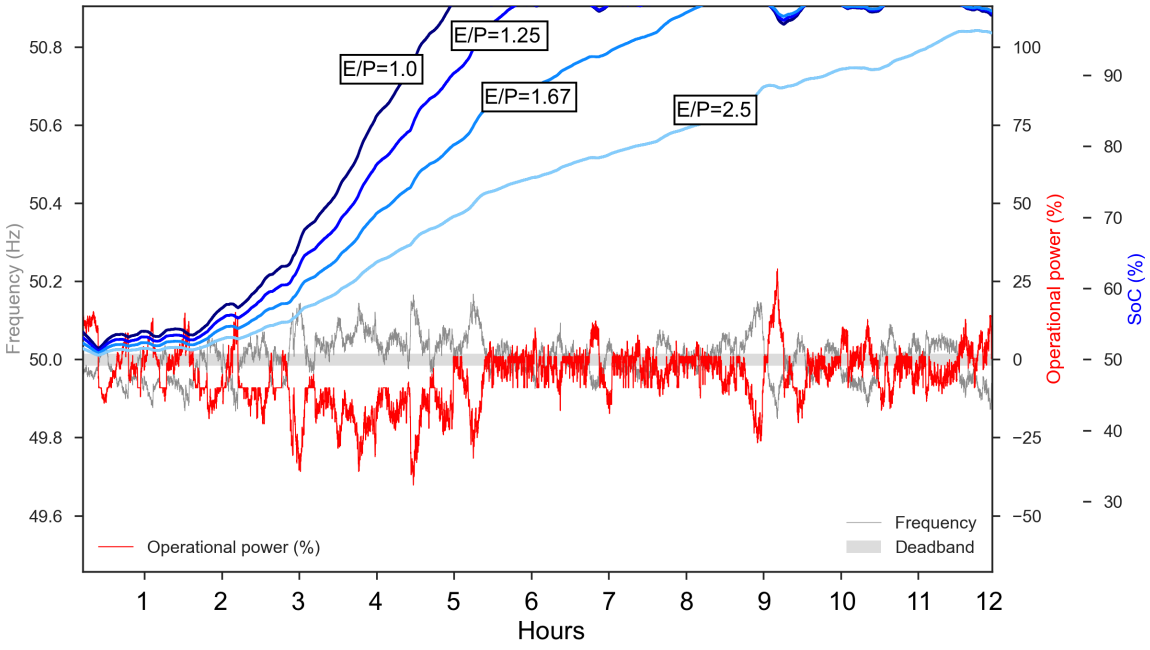

Figure 76: BESSs with different energy/power ratios achieving $SoC_{ddsp} = 100\%$. The pre-RZM recharge periods are different in each case (100% round-trip efficiency is assumed)

As shown in Figure 77, there exists a near-linear relationship between the BESS energy/power ratio and the recharge period. This relationship was validated over different periods of frequency recordings over the year. Based on the energy/power ratio of a BESS, the recharge period per 1% increase in SoC could be derived using

the regression equation given in Figure 77. This is used to estimate the recharge period required to achieve the SoC_{ddsp}, every day ahead.

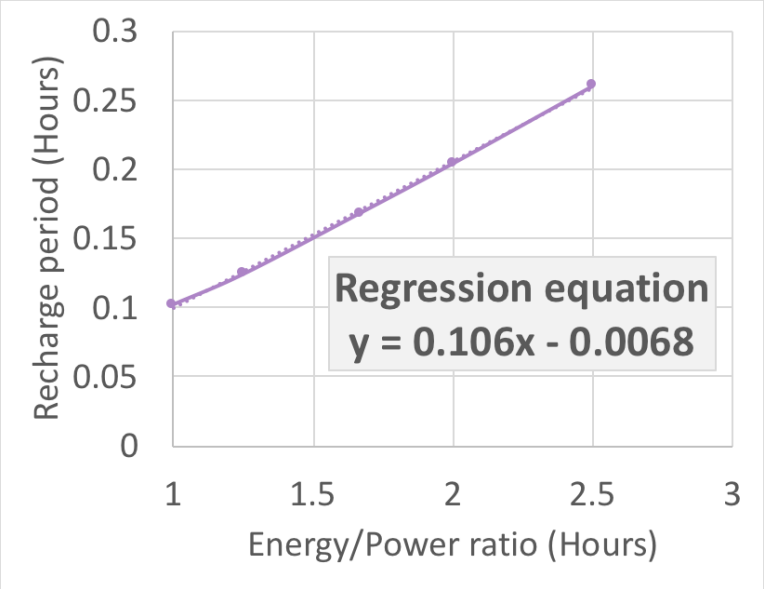

Figure 77: Linear relationship between energy/power ratio of the BESS and recharge period

6.4 Economic analysis

A consumer BESS participating in EFR and RZM DR programs, adopting the data-driven SoC management, completes a certain number of discharge cycles over a period of time in its life. Further, participation in these DR programs help the BESS owners earn revenue to recover the costs and make profits. The aspects of BESS life estimation and project viability assessment are discussed in this section.

6.4.1 BESS life estimation

Battery life is the maximum number of discharge cycles it can survive, given that the temperature is regulated. Discharge cycles are calculated based on different methods by different battery manufacturers. The methods also vary based on the battery chemistries. A simple method recommended by Apple Inc. [176] is adopted for battery discharge cycle calculation here.

In this method, a 100% discharge of a fully charged battery accounts to one discharge cycle. If the discharge occurs with intermittent charging as shown in Figure 78.

$$Total\ discharge = D1 + D2 + D3 = (100 - 60) + (80 - 0) + (70 - 40) = 150\%$$

$$Number\ of\ discharge\ cycles = \frac{150\%}{100\%} = 1.5$$

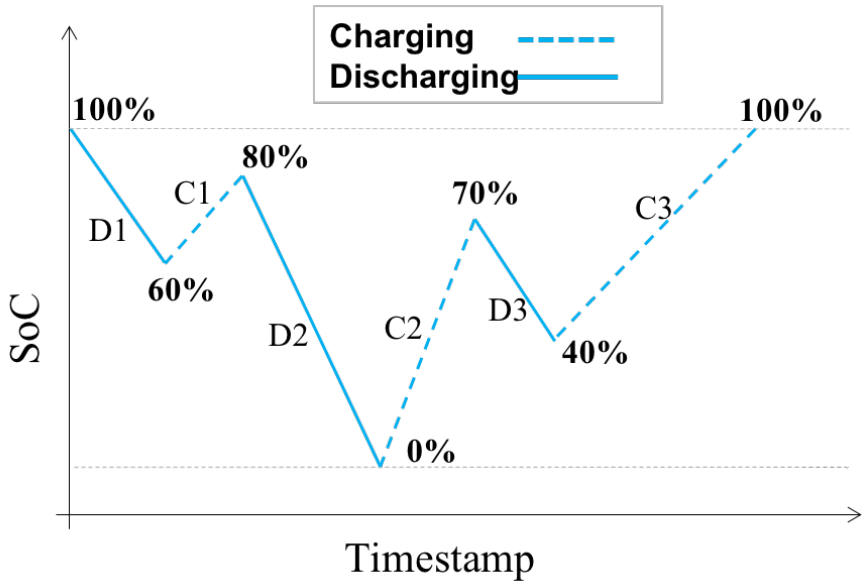

Figure 78: Battery discharge with intermittent charging

6.4.2 Project viability assessment

The viability of the BESS participating in EFR and RZM programs is estimated in terms of net present value (NPV) [108]. The initial capital *investment, cash inflow* through EFR and RZM participation and *cash outflow* for operation and maintenance of the BESS are used to evaluate the project viability.

$$NPV = \sum_{t=1}^{T} \frac{Cash\ inflow - Cash\ outflow}{(1+r)^t} - Investment \quad (6.6)$$

where T is the battery life and r is the discount rate to take the risk and inflation into account. A positive NPV shows that the project is viable.

6.5 Case study

In this section, a case study is performed to simulate the data-driven SoC strategy on a university building in the UK. The simulation methodology and results are discussed in Section 6.5.1. Following the simulation, an economic analysis of the BESS installation is performed in Section 6.5.2.

6.5.1 Simulation of the data-driven SoC management strategy

Participation of a university building BESS in the EFR and RZM programs based on the data-driven SoC management strategy is simulated for different scenarios using a set of codes written in the Python environment. As part of the simulation, BESS design parameters selected based on the suitability for the building site are discussed in Section 6.5.1.1. DR program specifications that dictate the BESS operation are discussed in Section 6.5.1.2. A supervised ML based load forecast model is developed in Section 6.5.1.3. Different scenarios of data-driven set-point SoC and recharge periods are derived and simulated for a sensitivity analysis in Section 6.5.1.4. The sensitivity analysis is performed to assess the robustness of the developed SoC management strategy.

6.5.1.1 BESS design parameters

Based on the annual meter data, the building peak load during the 3 hour RZM period (16:00 to 19:00) is 0.824 MW and the peak energy consumption for this period is, E_{backup_peak} = 2.17 MWh. The lithium iron magnesium phosphate ($LiFeMgPO_4$) based battery chemistry is ideal for large scale grid connected BESS applications [177], and this is considered for the case study. Rated power of the BESS is taken as 1.65 MW, twice that of the recorded building peak load. Energy capacity of the BESS, $E_{battery}$ in MWh should be sufficient to back up the E_{backup_peak} recorded. Since energy capacity is the main determinant of the BESS's capital cost, different possible values for $E_{battery}$ are considered in this study: 1.25, 1.5 and 1.75 times the E_{backup_peak}. These are used in the simulations and sensitivity analysis in Section

6.5.1.4. The discharge and recharge efficiencies of the BESS are 95% each, giving a roundtrip efficiency of ~90%. Round-trip efficiency is the product of recharge and discharge efficiencies. In the real-world implementation, monitoring and control of the BESS is dictated by the battery management system (BMS). However, the BMS design aspects are not included in the simulation. All the BESS design parameters used in the simulation are summarised in Table 13.

Table 13: BESS design parameters used in the simulation

Rated power	1.65 MW
Energy capacities considered	$1.25 * E_{backup\_peak}$ $1.50 * E_{backup\_peak}$ $1.75 * E_{backup\_peak}$
Charging efficiency	95%
Discharging efficiency	95%
Round-trip efficiency	~90%

6.5.1.2 DR program specifications

The DR programs were discussed in detail in Section 6.2. This section summarises the DR program specifications used in the simulation.

As part of the EFR program participation, the BESS is simulated to provide high and low frequency responses (operational power import and export, respectively) within 1 second when the grid frequency is outside a narrow deadband. When grid frequency is within the deadband, SoC management is allowed using a slow operational ramp until the operational power limits are reached. Annual GB system frequency data from 2015 (1 second resolutions) are used for the simulations. The EFR program specifications used in the simulation are listed in Table 14.

Since the BESS also participates in the RZM program from 16:00 to 19:00 hrs, it provides backup energy to the university building and does not deliver frequency response during this period. RZM program specifications used in the simulation

include the start and stop times of building load curtailment. These are also listed in Table 14.

Table 14: DR program specifications used in the simulation

EFR program (high and low frequency response) specifications	
Low frequency response	<49.985 Hz and >=49.500 Hz
High frequency response	>50.015 Hz and <=50.500 Hz
Deadband	50.015 Hz - 49.985 Hz
Response time	1 seconds
Within deadband, operational ramp	0.02 MW/s
Within deadband, operational power limit	0.18 MW
Slow operational ramp	0.18 MW
RZM program (evening demand peak shaving) specifications	
Load curtailment start time	16:00 hrs
Load curtailment stop time	19:00 hrs

6.5.1.3 Building load estimation modelling

As part of the case study, load estimation modelling for the university building is performed based on the data-driven approach developed in Chapter 4. Annual meter data are collected at half-hourly intervals from the university building. The building load variations to time-of-day, day-of-week and month-of-year variables are shown in Figure 79. The load variations to temporal predictors resemble that of a typical office building as elaborated in Chapter 4 Section 4.2.3.1.

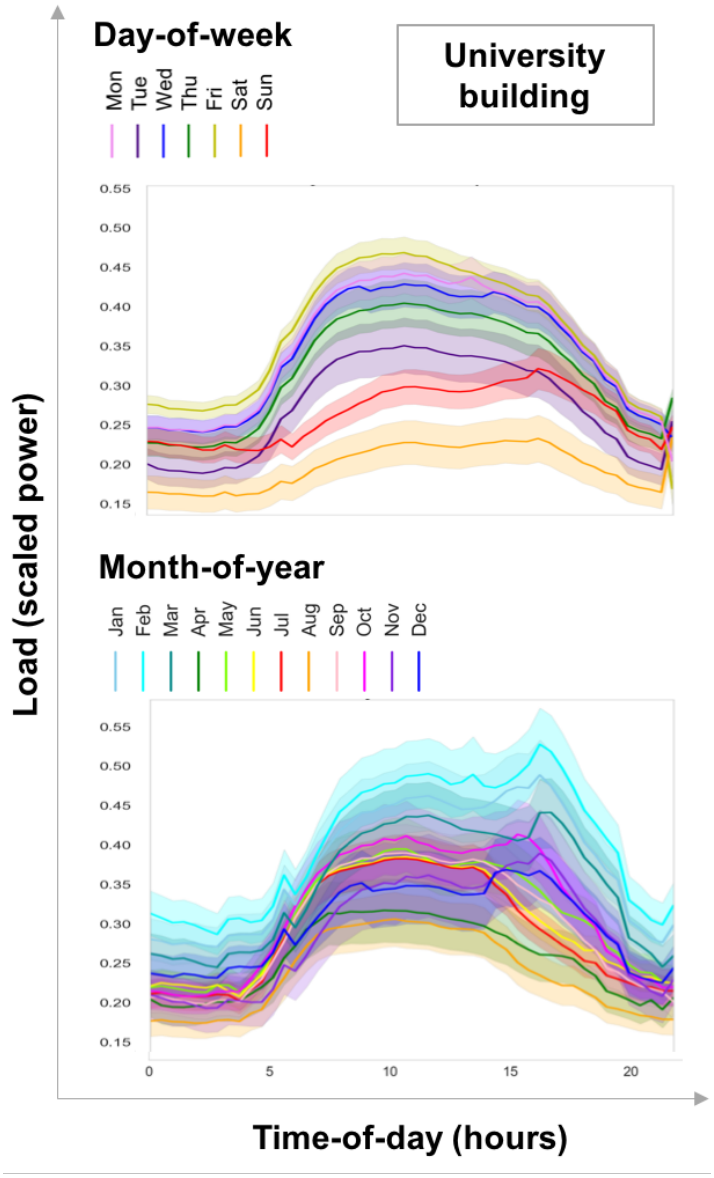

Figure 79: University building load variations (band shows 95% confidence interval)

The gradient boosted trees (GBT) based day-ahead supervised ML regression models were observed to provide the best predictive performances and the lowest model calibration times for multiple large consumer buildings in Chapters 4 – 5. For this reason, GBT algorithm is preferred for developing a day-ahead load estimation model for the university building. Time-of-day and day-of-week variables are extracted from the meter data timestamp and used as features for the model. Weather variables are not collected. Testing set size is set to 1 day, since the load

forecasts should be available day-ahead. The training set precedes the testing set without any gaps between them. This approach is motivated by the experimentation performed in Chapter 4 Section 4.3.1.6.

The training set size is selected using the learning curves method discussed in Chapter 4 Section 4.3.1.5. Training set sizes incremented in days up to 12 days are used for building different GBT models. For each model, 50 training-testing sets are randomly sampled using forward sliding window for cross-validation. The training and testing errors are measured using the mean absolute error (MAE) metric. The learning curves plot in Figure 80 shows the average training and testing errors for each of the GBT models. Based on this plot, the training set size is selected as 7 days since there is no decline in the average testing errors with further increase in training set size.

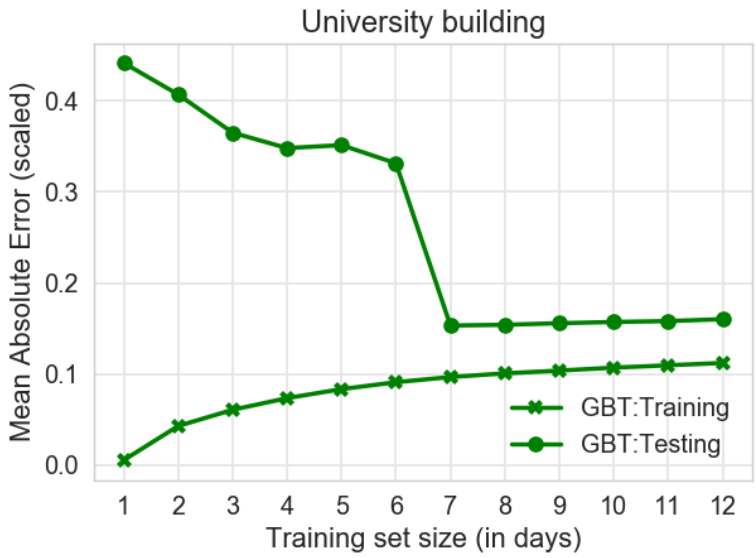

Figure 80: Learning curves to determine training set size for the university building dataset using gradient boosted trees (GBT) algorithm

The GBT model with testing set size of 1 day and training set size of 7 days, is calibrated using random-search hyperparameter optimisation (HPO) described in Chapter 4 Section 4.3.3.1. The 50 training-testing sets randomly sampled previously are equally split into a development set and an evaluation set; giving 25 to each set.

One stage of model calibration is performed on the development set using 75 hyperparameter sets randomly selected from the trial values listed in Chapter 4 Table 6. The model calibration process using GBT algorithm takes ~18 minutes. The best set of hyperparameters giving the lowest average MAE in the development set are: *number of base models* (500), *loss function* (LAD), *maximum depth of a tree* (49), *minimum samples to split* (5). The calibrated model is further validated in the evaluation set, which gives a mean absolute percentage error (MAPE) of 4.7%. This is set as the benchmark performance prior to deployment and the model is used for day-ahead building load estimation in the BESS operations.

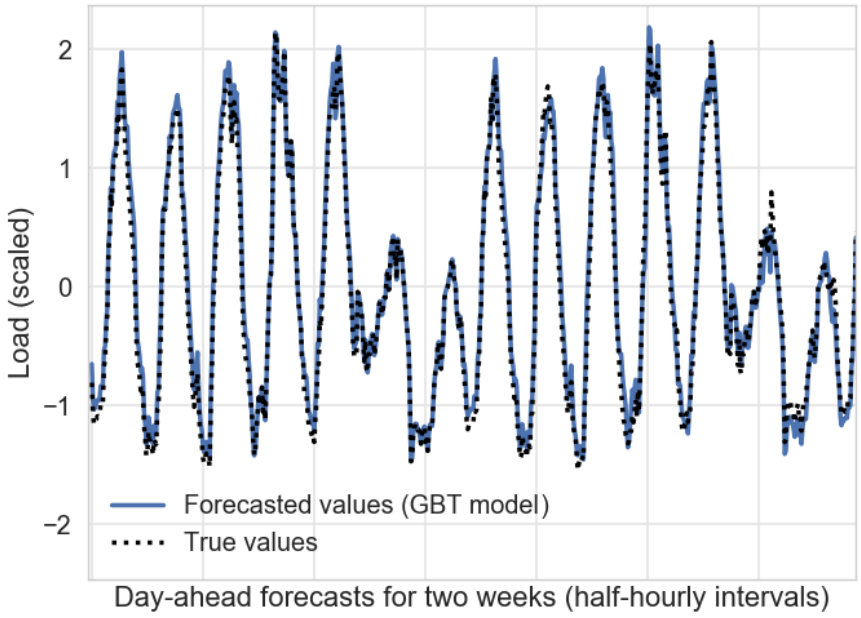

Figure 81: Day-ahead forecasted and true values of the university building load

6.5.1.4 Simulation process, sensitivity analysis and outcome

The half-hourly intervals of building load data (true and forecasted) from the 25 testing days in the evaluation set considered in the previous section are used for the simulation purpose. These data are resampled to 1 seconds to match with that of the frequency data for 25 days from the 2015 dataset. The days are not matched since the frequency variations are random throughout the year.

As part of the simulation, the building load values (true and forecasted) in power units are converted to energy units. Using the daily backup energy E_{backup} (true and forecasted) for the 3 hour RZM period (16:00 to 19:00 hrs) in equation (6.5) the daily data-driven set-point SoC (SoC_{ddsp}) values are derived. Based on the daily SoC_{ddsp} values, the BESS energy/power ratios and the corresponding regression equations (given in Figure 77) for 90% round-trip efficiency, the daily recharge periods are also calculated. A sensitivity analysis is performed to assess the extent up to which forecasting errors can be tolerated. For this purpose, errors of 5%, 10% and 15% are added and subtracted to the true $E_{backup.}$ Subsequently, the corresponding daily SoC_{ddsp} and recharge periods are calculated. In summary, the different scenarios of SoC_{ddsp} and recharge periods considered for the simulation are: *true*, *true+error* (error = +5%,+10%,+15%,-5%,-10%,-15%) and *forecast*. The simulations are run for different BESS energy capacities (as listed in Table 13).

Outputs of the simulations include number of discharge cycles, operational range of SoC, counts and durations of service disruptions, for different scenarios and BESS energy capacities. The number of discharge cycles are calculated based on the method described in Section 6.4.1. Operational range of SoC tracks the minimum and the maximum observed during the period of simulation (25 days in this case) Service disruptions are considered for the EFR window when the SoC reaches 0% or 100%, restricting the BESS to export energy (for providing low frequency response) or import energy (for providing high frequency response), respectively. Service disruptions of 2 seconds and above are accounted for. The simulation results for different scenarios and different BESS energy capacities are given in Table 15. The scenarios with service disruptions are marked red and their discharge cycles are not recorded (hence marked N/A).

Table 15: Simulation results using different scenarios; those with service disruption are marked red.

SoC_{ddsp} Scenarios	$E_{battery} = 1.75*(E_{backup\_peak})$				$E_{battery} = 1.50*(E_{backup\_peak})$				$E_{battery} = 1.25*(E_{backup\_peak})$			
	Cycles	SoC range	Service disruption		Cycles	SoC range	Service disruption		Cycles	SoC range	Service disruption	
			Counts	Duration (mins)			Counts	Duration (mins)			Counts	Duration (mins)
True + 15% error	13.72	81 - 9 %	0	0	N/A	86 - 0 %	3	0.03 – 9	N/A	95 - 0 %	3	0.03 - 10
True + 10% error	13.9	79 - 6 %	0	0	N/A	84 - 0 %	3	0.03 – 6	N/A	90 - 0 %	3	0.03 - 10
True + 5% error	13.64	78 - 6 %	0	0	15.71	81 - 4 %	0	0	19.08	93 - 1%	0	0
True.	13.79	77 - 4 %	0	0	16.35	82 - 3 %	0	0	N/A	91 - 0 %	3	0.03 - 8
True – 5% error	13.48	76 - 9 %	0	0	15.79	80 - 3 %	0	0	N/A	90 - 0 %	4	0.03 - 15
True – 10% error	13.72	74 - 4 %	0	0	16.14	79 - 4 %	0	0	N/A	86 - 0 %	4	0.03 - 15
True – 15% error	13.75	73 - 4 %	0	0	N/A	77 - 0 %	2	0.03 – 5	N/A	84 - 0 %	4	0.03 - 15
Forecast (MAPE = 4.7%)	13.95	81 - 2 %	0	0	16.01	80 - 3 %	0	0	N/A	95 - 0 %	5	0.03 - 17

Simulation of 25 days BESS operation shows that the likelihood of service disruptions is very high when the BESS energy capacity $E_{battery} = 1.25 * E_{backup\_peak}$. The SoC is observed to reach 0% in most scenarios. Even for the *true* scenario, 3 service disruptions with durations in the range of 0.03 – 17 mins are observed. Although large enough to backup the peak energy consumption during the RZM period, the BESS energy capacity is not sufficient to maintain reliable operation in the two DR programs simultaneously.

Service disruptions for a BESS with energy capacity $E_{battery} = 1.50 * E_{backup\_peak}$ occur only when the forecast errors are above +10% or below -15% of the true E_{backup}. The GBT based forecast model has a MAPE of 4.7% which is well inside these error bounds.

When $E_{battery} = 1.75 * E_{backup\_peak}$, service disruptions are not observed for any of the scenarios. Since errors up to +/- 15% are observed to be tolerated, a forecast model with larger errors may also be implemented in this case. Also, the number of discharge cycles are lower compared to small battery energy capacities, giving extra lifetime for the BESS. However, the large BESS energy capacity comes at an increased capital cost.

Based on the simulations, it may be concluded that, a BESS with energy capacity 1.5 times that of the peak energy backup (i.e. 3.25 MWh) is the most optimal for

installation in the university building site. This BESS is considered for the economic analysis in the next section. The GBT based day-ahead building load estimation model is reliable due to its lower MAPE. Figure 82 shows the operation of this BESS in EFR and RZM programs based on the data-driven SoC management strategy for 3 consecutive days out of the 25 days considered.

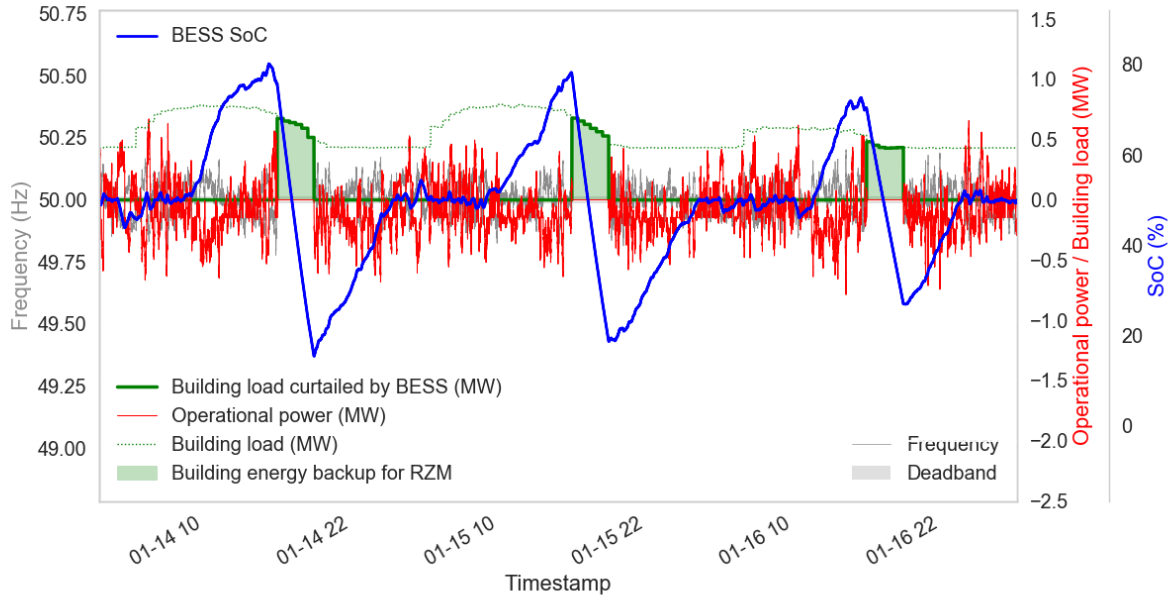

Figure 82: Reliable EFR and RZM operation of a 1.65 MW 3.25 MWh BESS for three consecutive days using the day-ahead forecasts of SoC_{ddsp} and recharge periods, without any service disruptions (timestamp format is: month-day hour)

6.5.2 Economic analysis of the university building BESS

Figure 83 shows the battery cycles in a lifetime versus depth of discharge for the $LiFeMgPO_4$ chemistry. Depth of discharge refers to how deeply the BESS is discharged. The plot is used to calculate battery life based on the number of cycles and depth of discharge.

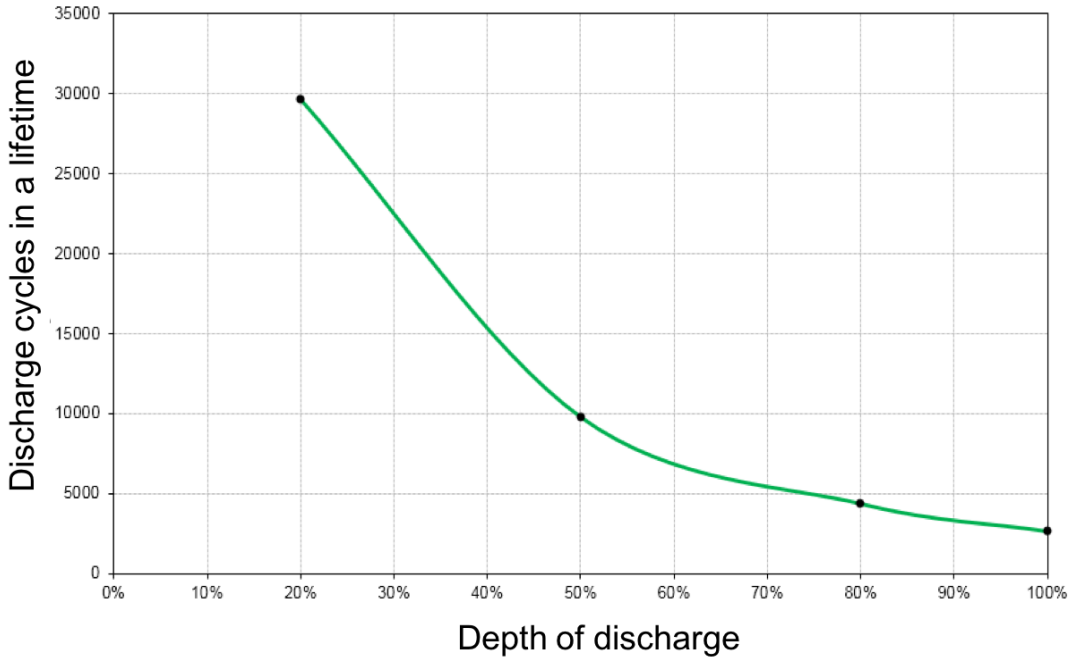

Figure 83: Discharge cycles in a lifetime versus depth of discharge plot for the $LiFeMgPO_4$ battery; adapted from [177].

Based on the outputs of the simulation performed in the previous section and using the curve plotted in Figure 83, the lifetime of a 1.65 MW 3.25 MWh BESS participating in EFR and RZM programs is estimated for true and forecast scenarios in Table 16. The calculated battery cycles for these scenarios are approximations based on the BESS operation simulated on randomly sampled 25 days which are assumed to be representative of all other days. The BESS lifetime estimated based on the true scenario is 19.27 years and that based on the forecast scenario is 20.11 years, which are not substantially different.

Table 16: Lifetime estimation of the 1.65 MW 3.25 MWh BESS for true and forecast scenarios

	True	Forecast
Cycles in 25 days	16.35	16.01
Cycles per day	0.65	0.64
SoC range	82 - 3%	80 - 3%
Depth of discharge	79%.	77%.
Cycles in lifetime	~4600	~4700
Lifetime (years)	19.27	20.11

A conservative lifetime of 19 years is considered for the project viability analysis. The capital cost of the BESS is assumed to be £400/kWh. For a 3.25 MWh BESS this would total to £1.30 million. The annual maintenance cost is assumed to be £1000. The availability payment for EFR program is £9.44/MW/hour. Subtracting the RZM period (16:00 to 19:00), this would mean a participation of 21 hours per day or 7665 hours per year. Hence the annual EFR revenue from the 1.65 MW BESS is £119390. The revenue from RZM participation is the avoided red zone consumption price of £7.632/kWh. Based on the annual load during the RZM period, the avoided red zone consumption is estimated to be 378 MWh. Hence the annual revenue from RZM participation would be £2.88 million. Based on equation (6.6) the NPV is calculated to be £34.99 million. Since it is a positive value, the BESS installation in the university building site is considered profitable.

6.6 Discussion

A critical discussion on the assumptions made in the study and applicability of the developed data-driven SoC management strategy follows.

The DR programs considered in this chapter consist of a frequency response program (EFR) and an evening demand peak shaving program (RZM) in the UK. The novel data-driven SoC management strategy is developed around the

framework of these programs. However, the specifications of these DR programs are susceptible to changes in the future. For example, the EFR program in the UK was introduced to gauge the interest for dynamic frequency response from BESS developers. In a future version of this DR program, if the deadband operational ramp and power limits are decreased, the BESS would take longer time to manage its SoC. This would extend the daily recharge periods prior to the evening peak period, leading to more service disruptions. On the positive side, if the existing deadband operational ramp and limits are increased, the recharge periods will be smaller and service disruptions will be minimised. As a result, the data-driven SoC management strategy becomes more reliable.

The discharge cycle calculation is performed based on a method discussed in Section 6.4.1. It is possible that the $LiFeMgPO_4$ battery manufacturer may have followed a different method for this purpose. The discharge cycles in a lifetime vs depth of discharge curve is used for BESS life estimation based on the assumption that the calculated cycles are accurate. Further, the influence of factors such as temperature on the BESS life has not been taken into account while performing the simulations. This is considered as a future work.

Simulations of the data-driven SoC management strategy are performed on 25 randomly selected days of the current year. The operational range of SoC observed for different scenarios based on these days are merely indicative. In the future, if longer term building load meter data are available, more number of days may be used for the simulation.

Further, the BESS rated power and energy capacity are selected based on the peak building energy consumption in the RZM period from the collected annual meter data. The simulation proposes a lifetime of ~19 years, in the course of which, the building energy consumption may increase. This may add to the upgradation costs of the BESS, which are not taken into account in the economic analyses performed.

Despite these assumptions, the novel SoC management strategy developed in this chapter is applicable for reliable operation of a consumer BESS in multiple DR programs. Simulations and sensitivity analysis performed as part of the case study on a university building demonstrates the robustness of this approach. Simulation results show the trade-off between forecast errors and energy capacities. With increase in errors in the data-driven set-point SoC, larger energy capacities are required, adding to the BESS's capital cost. This provides conclusive evidence on the fact that improvement in the predictive performance of the building load estimation model can help reduce the capital cost.

6.7 Conclusions

Due to the limited energy capacity of a BESS, SoC management is an important aspect for its reliable operation in DR programs. Different SoC management strategies based on deadband, real-time offset, scheduling and EVs were explored in the previous literature. With increasing interest in consumer BESS (or behind-the-meter) projects, it is essential to adopt new SoC management strategies that suit specific scenarios of DR participation.

A data-driven SoC management strategy was developed for a large consumer building BESS participating in DR programs related to frequency response and peak shaving in the UK electricity market. The strategy utilises the deadband SoC management allowance of an existing frequency response program (EFR) to ensure sufficient SoC for participation in the peak shaving program (RZM). This is guided by a set-point SoC that is estimated one day in advance using a data-driven load estimation model. This data-driven set-point SoC helps estimate a recharge period every day in advance, provided as input to the BESS controller.

The data-driven SoC management using the deadband allowance avoids the need for scheduling BESS recharge in the pre-RZM period. A scheduling BESS recharge approach would result in revenue loss for the consumer since no EFR participation

is possible during that period. Further, if a large number of consumer BESS units perform the scheduling BESS recharge in the pre-RZM period, grid balance would be in jeopardy.

The SoC management strategy was demonstrated on a university building with 1.65 MW and 3.25 MWh lithium iron magnesium phosphate BESS participating in the EFR and RZM DR programs. Based on the number of cycles, the BESS was shown to have a conservative lifetime of ~19 years. The calculated NPV was positive and hence the project is considered viable. However, it is expected that, for an earlier return on investment, a higher revenue could be offered to the university building BESS for participation in the EFR (or EFR like) DR programs.

To the best of our knowledge, data-driven SoC management strategies are seldom studied in the case of BESS based DR participation. In this chapter, the application of data-driven modelling in the reliable operation of a BESS is shown to add value to the consumer as well as the grid. The developed method may motivate further research on data-driven SoC management that would enable reliable operation of BESS in grid balancing services of today and the future.

Chapter 7

Emissions reduction potential of large consumer buildings participating in an emissions-based DR program

7.1 Introduction

Chapters 4 – 6 explored the development of data-driven building load estimation models and their applications in different demand response (DR) operations. For data-driven modelling purposes, consumer smart meter data and/or weather data were used. This chapter discusses the use of smart grid data such as grid emissions intensity for a proposed emissions-based DR program and assesses the emissions reduction potential for participating large building consumers. The use of data-driven pipelines in utilising the vast amounts of data available from smart grids and delivering value to the consumers and environment is highlighted in this chapter.

7.1.1 Background

Increasing the share of renewables in the global electricity fuel mix is significant in the wake of emission reduction targets set by economies around the world. Although the cost of renewables is declining, given the intermittency of resources such as wind, the incremental increase in electricity demand will have to be met by marginal generators that run on gas or coal. A marginal generator is likely to not run if the demand could be curtailed in short notice. Hence, the transition to a low emission grid could be accelerated through mechanisms such as DR that involve the consumer loads.

Conventional DR programs being rolled out by the system operators or utilities require loads to be curtailed or shifted in time, based on dynamic price, reliability signals (such as grid frequency) or notification provided to the consumer site. However, the load curtailment or shift triggered by these conventional DR programs need not necessarily reduce the grid emissions from marginal generators. For instance, dynamic pricing-based DR programs are designed based on market prices that do not reflect the environmental cost of emissions, i.e. peak prices are only indicative of high demands and not high grid emissions. Consequently, curtailment of load as a response to high prices does not guarantee reduction in grid emissions. As another example it can be stated that, frequency-based DR programs are designed to care for the electricity system health and not the environmental health.

Grid emissions intensity is the volume/weight of greenhouse gas emissions per unit of electricity generated. As a result of smart grid development in different electricity markets, grid emissions intensity data (referred to as grid emissions data for simplicity) are publicly available in real-time. It is possible to utilise the existing DR mechanism and available flexibility on the consumer side to trigger load curtailment based on these data. In contrast to conventional DR programs, such an emissions-based DR program will provide consumers the choice of energy based on the emissions attached to every unit generated. It has been reported that, approximately 60% of the US corporations already have emission reduction targets in place [178]. The emissions-based DR program can help such corporations to reduce their annual carbon emissions. With increased participation in this DR program, the transition to a low emissions grid could be accelerated.

7.1.2 Contributions

Conventional DR programs in electricity markets focus on grid balancing and not on emissions reduction. This study contributes to the concept of an emissions-based DR program targeted at grid emissions reduction through load curtailment of large

consumers. Although beneficial, emissions-based DR programs are non-existent in any of the electricity markets.

One of the primary requirements to realise an emission-based DR program is the availability of real-time grid emissions data. The grid emissions intensity data used in this study is derived using a method that estimates them from locational marginal prices (referred to as LEEM 2.0 emissions estimation model) and is accessed from a continuously updated open database. To the best of our knowledge, this data has not been used to assess the emissions reduction potential of large consumer buildings.

The study was conducted in the United States (US) for certain system operator regions. The share of flexible loads in large consumer buildings already participating in conventional DR programs were quantified from the energy consumption data collated through DR aggregators in the US. The load flexibility in large consumer buildings was further validated based on the data from relevant literature. The proposed emissions-based DR program is designed around the emissions intensity data mentioned earlier. Emissions reduction potential of large consumer buildings in the selected system operator regions are assessed for multiple scenarios of load flexibility and durations of curtailment annually. These results are also compared with the emissions reduction possible from conventional DR programs. The study demonstrates the emissions reduction potential for large consumers through participation in the proposed emissions-based DR program.

7.1.3 Layout

The remaining part of this chapter is structured as follows. Methods adopted towards estimating the emissions reduction potential of large consumer buildings are discussed in Section 7.2. The results for different scenarios are discussed in Section 7.3. Implementation aspects of the emissions-based DR and conclusion of the chapter are given in Section 7.4.

7.2 Methods

This section elaborates on the methods adopted in the study. The model used to derive the grid emissions data is described in Section 7.2.1. Available flexibility in large consumer buildings are quantified based on field data and literature survey in Section 7.2.2. Based on this information, different scenarios of load flexibility and durations of curtailment are developed towards assessing the emissions reduction potential of large consumers in 7.2.3.

7.2.1 LEEM 2.0 grid emissions estimation model

Emissions from grid connected generation depend on many factors such as type of fuel used, efficiency of the generator, pollution control devices used and location of the power plant. Such information for marginal generation is confidential due to regulatory restrictions and hence inaccessible. A method developed by Carter et al. [179] and improvised by Rogers et al. [180] estimates emissions from marginal generation based on locational marginal prices (LMPs) that are available publicly. LMP is the wholesale market price for the marginal unit of electricity generated at a given location and is derived as a function of the cost of generation, transmission constraints and system losses. The improvised method referred to as LMP emission estimation method (LEEM 2.0) utilises LMPs to identify pollutant emission factors of marginal generators based on the fuel type for a given location and time [180]. Historic and real-time five minutes resolution grid emissions intensity data based on LEEM 2.0 method is publicly available for all the system operator regions in the US, here [181]. A snapshot of this is shown in Figure 84. For the purpose of this study, historic marginal carbon emissions intensity data at 5 minutes resolution are collated from this database for three system operator regions viz. California Independent System Operator (CAISO), Electric Reliability Council of Texas (ERCOT) and Pennsylvania-New Jersey-Maryland (PJM) interconnection. These regions are represented in Figure 85. The proposed emissions-based DR program utilises these data for operations.

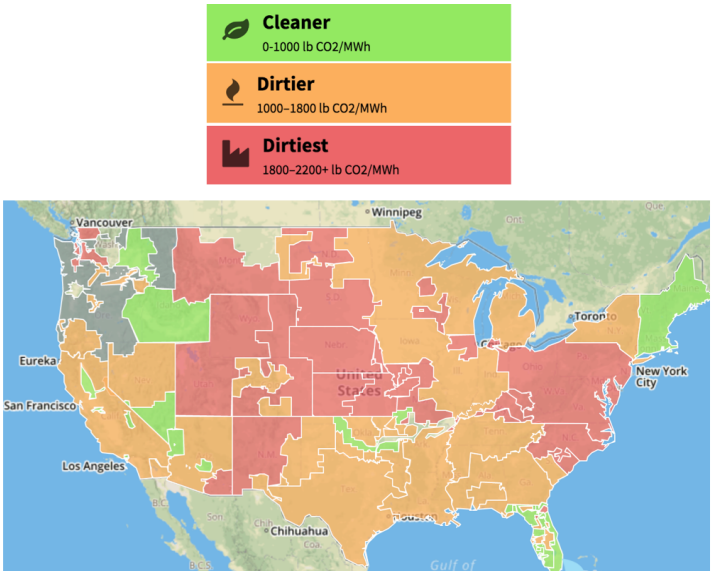

Figure 84: Real-time grid carbon emissions intensity map for the US [181]

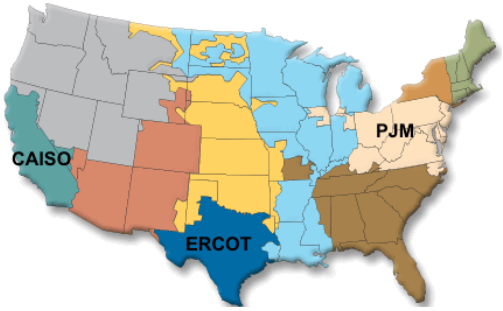

Figure 85: System operator regions in the US considered in the study [182]

7.2.2 Available flexibility in large consumer buildings

In order to assess the emissions reduction potential of large consumer buildings, it is important to quantify the available flexibility in these buildings first. This is performed through data collected from DR aggregators with a portfolio of large consumers and later validated with information in the relevant literature.

Data regarding large consumer buildings participating in conventional DR programs in the US were collected from DR aggregators in the three system

operator regions: CAISO, ERCOT and PJM. These data include annual meter data at 15 minutes resolution including load curtailment events, durations of the curtailment, estimates of the baseline loads as reference to what the building would have consumed in the absence of the curtailment events and control strategies adopted for curtailment by the respective buildings.

Figure 86 shows the load curtailment provided by four different retail buildings in the CAISO region. The dotted line shows the baseline load of the buildings. These buildings were reported to have used one of the following strategies for curtailment of their HVAC systems: duty cycling of rooftop units or chiller ramp down. The available flexibility ranges between 29% and 51% of the total load in each building. The curtailment was provided for a duration of 240 to 495 minutes (4 to 8.25 hours). Signs of rebound due to the load curtailment is evident in two retail buildings out of the four. However, the rebound is observed to be below 50% of the curtailed load.

Figure 87 shows the load curtailment in: 1) a retail and an office building in the ERCOT region and 2) an industrial building in the PJM region. The retail building provided 44% load curtailment for 150 minutes. This was based on a temperature offset strategy on the HVAC load performed for system demand peak shaving. Although it shows that the office building provided only 18% load curtailment, this was calculated using a baseline that is evidently underestimated. The duration of this emergency response DR event was 90 minutes. Minor rebounds are observed for the retail and office buildings. The industrial load in PJM was 100% curtailed for 60 minutes.

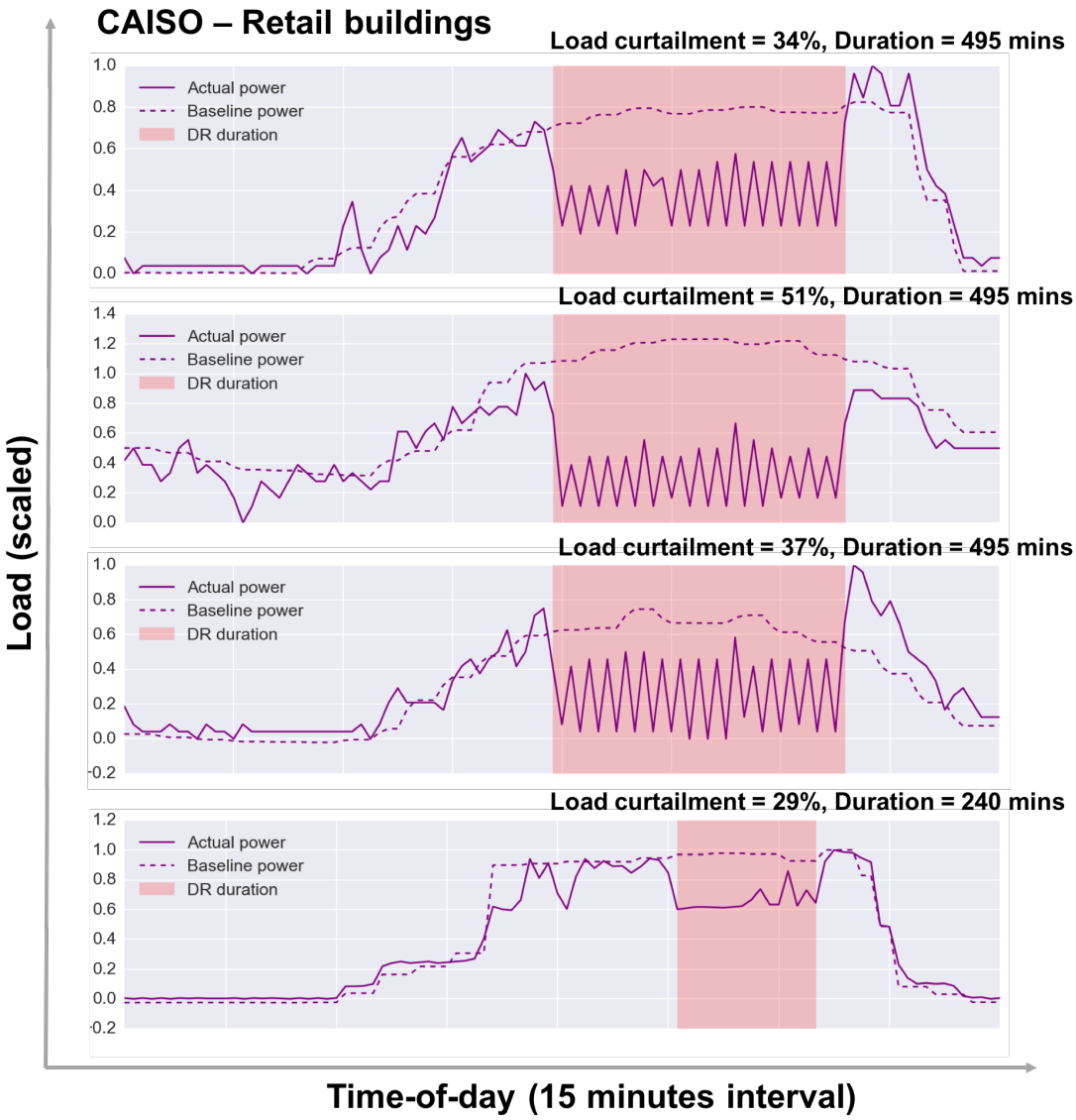

Figure 86: Retail buildings in CAISO participating in conventional DR programs

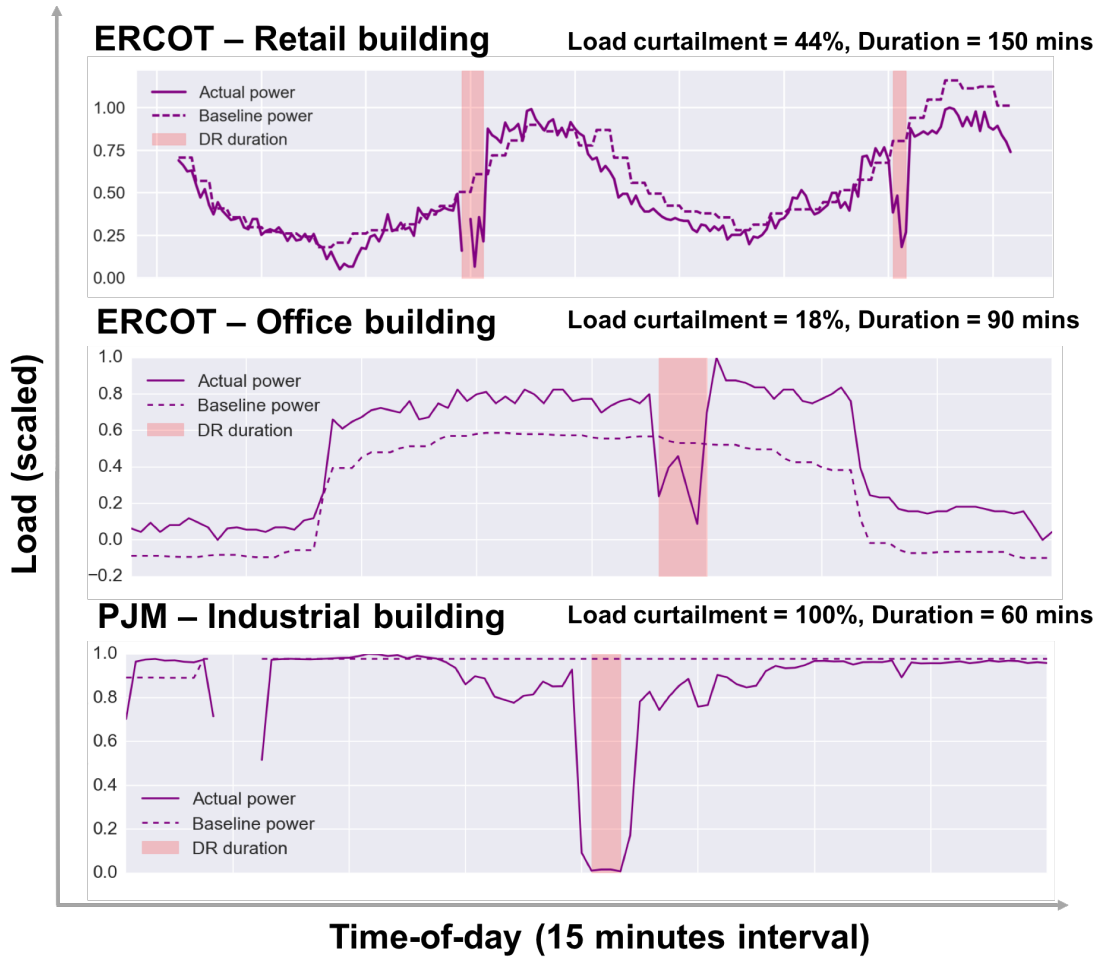

Figure 87: Commercial buildings in ERCOT and industrial building in PJM participating in conventional DR programs

Analysis of load curtailment events of large consumers participating in conventional DR programs in the US shows load flexibility up to 51% in commercial and 100% in industrial buildings. Load curtailment durations ranged between 1 and 8.25 hours per DR event. In addition to these field data collected from DR aggregators, diverse literature based on academic as well as industrial studies were reviewed to validate the available load flexibility. These are summarised in Table 17. Commercial load flexibility up to 56% and curtailment duration up to 6 hours per event are observed.

Table 17: DR strategies and availability of flexibility in large consumer buildings in the US

Building type	Building location	DR strategy	Load curtailment	Response duration	Source
Retail	California	RTU based	5-15%		[183]
Commercial	California	Global temperature control 0-3°C	10-16%		[164]
Large commercial	California	Control of variable air volumes (VAV), fans, lighting	19%	20-45 mins	[163]
Small commercial	California	Global temperature control	16-52%	20-45 mins	[163]
Supermarket	California	Temperature control 0-2°C	2-18%		[162]
Supermarket	California	35% lights reduction, anti-sweat door heater off	7-16%		[162]
Datacentre	California	Duct static pressure reduction from 1" to 0.5"; Hallway lights off	16-17%		[162]
Small office	California	Temperature set-point increase by 2°C	3-27%		[162]
Office, Laboratory	California	33-50% lighting off; package units with variable frequency drive (VFD) selective switch-off	7-56%		[162]
Large office	California	Temperature set-point increase by 2°C	4-14%		[162]
Large office	California	Temperature control +/- 2°C	4-30%		[162]
Laboratory	California	50% supply fans off; one air handling unit (AHU) off	7-22%		[162]
University	California	60% supply fans off; static pressure reset; heating & cooling valve position; economizer open	7-31%		[162]
Distribution centre	California	50% direct expansion (DX) package unit off	4-23%		[162]
Commercial	North Carolina	VFD-retrofitted HVAC supply fan;	28.5%	1 hour	[160]
C&I	New England	Lighting reduction in winter	16%	30-60mins	[184]
Retail	New England		20%	6 hours	[185]
Cold storage	Washington	Compressors & evaporators	21-32%		[186]
Datacentre	California	IT equipment load shed	25%	1 hour	[187]

7.2.3 Emissions reduction potential assessment

Available load flexibility and feasible duration of load curtailment in large consumer buildings were estimated in the previous section. Based on this information, in this section, different scenarios are designed for emissions reduction potential assessment of the large consumers who intend to participate in the emissions-based DR program.

The buildings considered for this assessment in each system operator regions are listed in Table 18. These buildings participate in conventional DR programs and are proven to have substantial load curtailment capabilities. The meter data shows the actual load and not the baseline load. Hence, these meter data include the load curtailment events as part of conventional DR programs. However, such events are small in number annually. Our analysis shows that conventional DR programs in the US last a maximum of 150 DR hours per year (i.e. the sum of load curtailment durations of all DR events per year).

Table 18: Study regions and type of buildings considered

System operator	Building
CAISO	3 retail
ERCOT	1 retail, 1 office
PJM	1 industrial

In addition to the conventional DR, these large consumer buildings can participate in the proposed emissions-based DR program where real-time grid emissions data for the respective system operator regions are available to them and a share of their flexible loads are ready to be curtailed. In the best interest of not interfering with the conventional DR program as well as not affecting the occupant comfort or business processes, multiple scenarios are considered based on: load curtailment of 1%, 5% and 10%; load shifting of 100% and 50%; DR hours of 1000, 2000 and 4000 per year.

Here, each scenario of participation is defined by the percentage of load curtailment, percentage of load shifting and DR hours.

Load curtailment of 1-10% from large consumer buildings is found to be very realistic, based on the analysis in Section 7.2.2. Load shifting takes into account the possible rebound due to curtailment of HVAC loads. In the case of industrial process, this would mean a 100% or 50% repetition of the operation that was curtailed. As observed earlier, in the case of conventional DR events, there are occasions where load shifting was negligible after a load curtailment event. Hence the values considered here are very conservative and on the safer side. DR hours considered for the emissions-based DR program is longer than the conventional DR programs. However, these are justified considering the smaller load curtailment in the range of 1-10%. Based on the grid emissions data, emissions duration curves are derived for each region. This is demonstrated in Figure 88 using 2015 data.

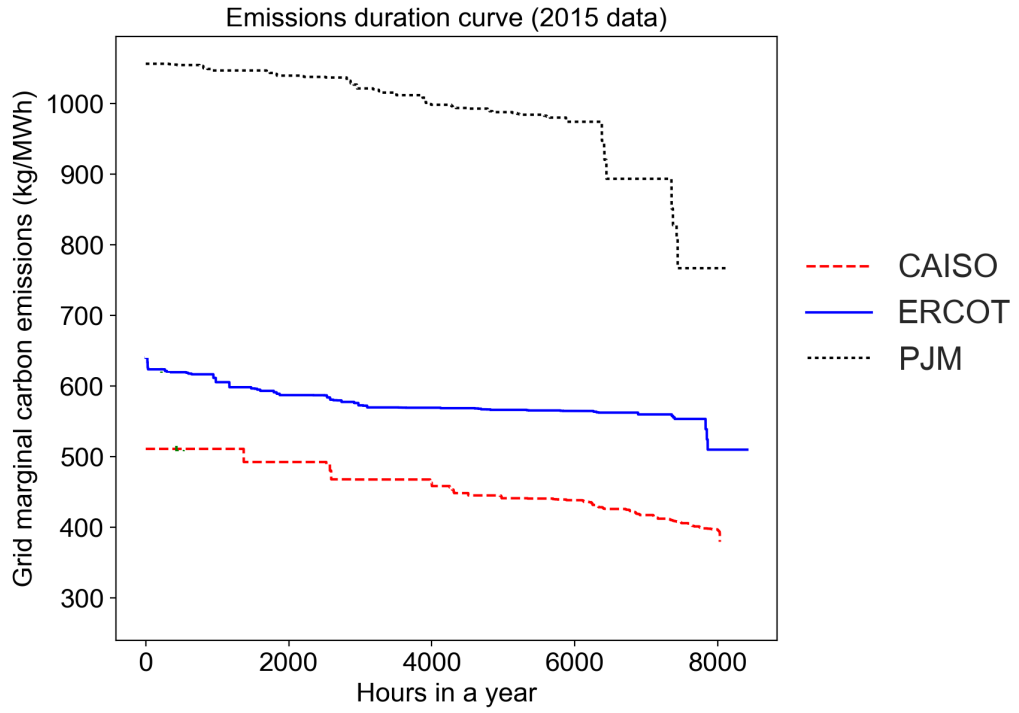

Figure 88: Grid marginal carbon emissions duration curves for different system operator regions considered in the study (2015 data resampled to 15 minutes)

Emission values are observed to vary across the regions, based on the share of polluting marginal generators. For each region, the emission duration curves are used to identify trigger values for achieving different load curtailment DR hours (1000, 2000 and 4000 hours per year). Here, trigger value refers to the grid marginal carbon emission intensity (kg/MWh) at which the building load needs to be curtailed so as to achieve a particular number of DR hours per year. As long as grid emissions are above these trigger values, the load curtailment continues.

The emissions reduction potential (%) for each building is calculated based on the general equation:

$$\frac{Annual\ avoided\ emissions\ through\ load\ curtailment\ (kg)}{Total\ emissions\ in\ the\ absence\ of\ load\ curtailment\ (kg)} \tag{7.1}$$

and the emissions reduction potential for each scenario of participation in the emissions-based DR is derived as:

$$\sum_{k=1}^{N} E_k . C_k \Big/ \sum_{i=1}^{M} E_i . C_i \tag{7.2}$$

where, 1) in the numerator: N is the number of 15 minute intervals over the year for which grid marginal carbon emission intensity is above the trigger value, E_k is the avoided energy consumption through load curtailment during each interval and C_k is the corresponding grid marginal carbon intensity; 2) in the denominator: M is the number of 15 minute intervals over one year (=35040), E_i is the energy consumption during each interval and C_i is the corresponding grid marginal carbon intensity. The emissions reduction potential for each building and scenario of participation are averaged for each system operator region.

As mentioned previously, the building meter data used for this analysis included load curtailment events due to participation in conventional DR programs. Using the baseline load data provided by the DR aggregator, the emissions reduction occurred through load curtailment triggered by conventional DR programs are also estimated for each building (the general equation above is used). Although

conventional DR doesn't intend to reduce grid carbon emissions, these estimates are compared with that of the emissions-based DR.

7.3 Results and discussion

Figure 89 shows the average emissions reduction potential of three retail buildings in the CAISO region, Figure 90 shows that of the retail and office buildings in the ERCOT region and Figure 91 shows that of the industrial building in the PJM region.

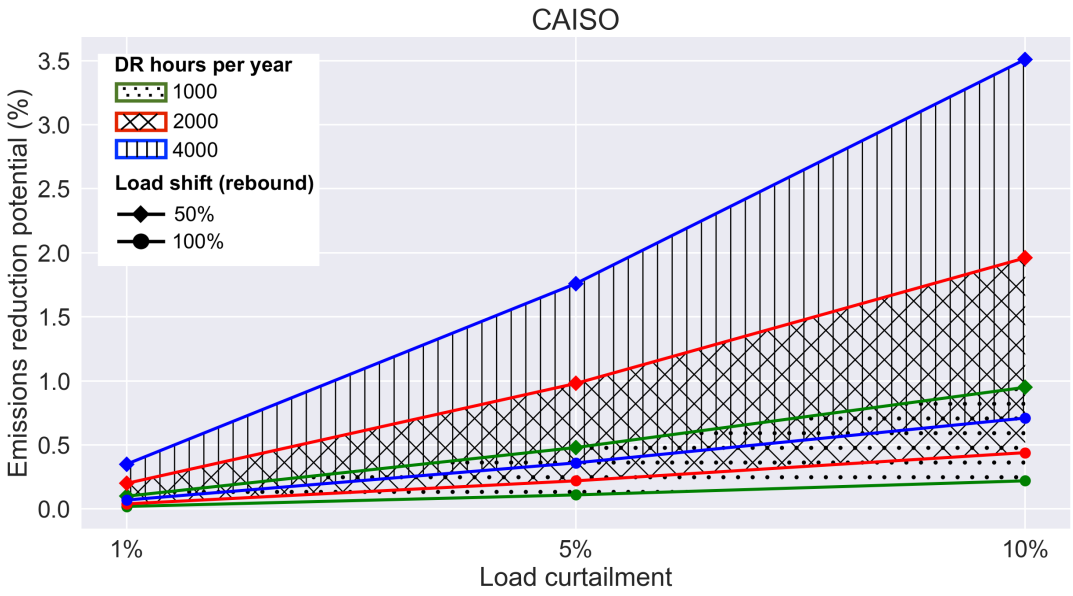

Figure 89: Emissions reduction potential for different scenarios is CAISO

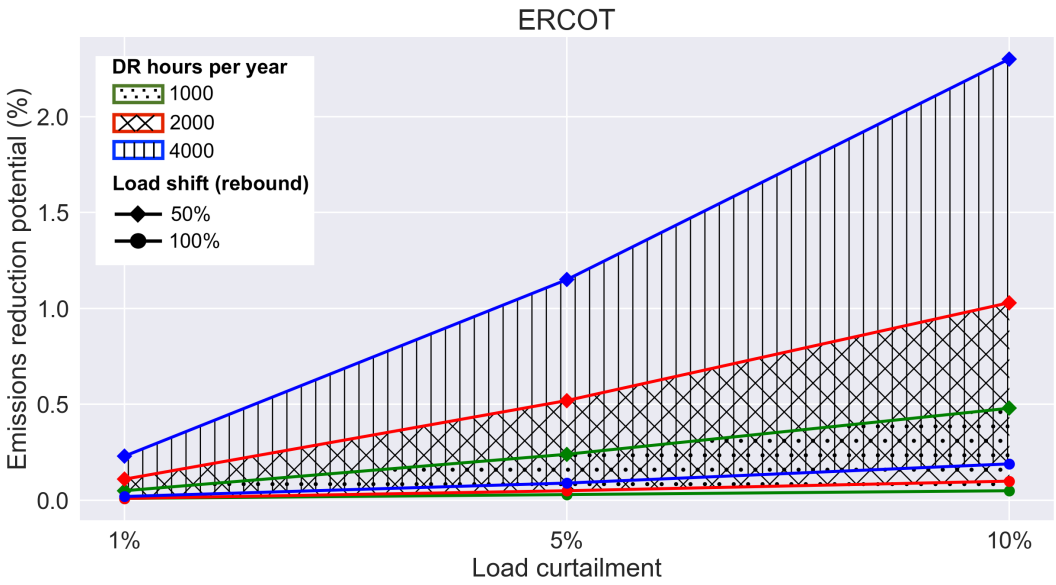

Figure 90: Emissions reduction potential for different scenarios is ERCOT

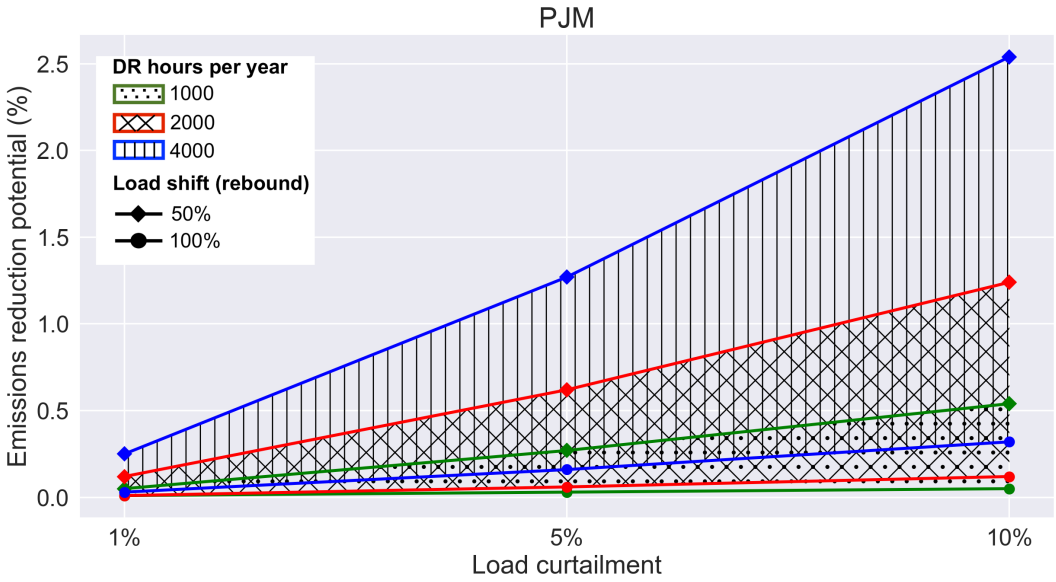

Figure 91: Emissions reduction potential for different scenarios is PJM

There is a linear increase in emissions reduction potential with increase in percentage of load curtailment. The gap between 50% and 100% load shift scenarios also widens with the increase in load curtailment. Based on the best case scenario of

10% load curtailment, 50% load shift and 4000 hours of DR, the emissions reduction potential for large consumer buildings in CAISO is 3.51%, in ERCOT is 2.30% and in PJM is 2.54%. Analysis of conventional DR events showed a maximum of 0.26% emissions reduction over all the buildings and regions considered. However, using the same DR controls for an emissions-based DR program, up to 3.77% emissions reduction could be unlocked from the large consumer buildings annually. A visual comparison is given in Figure 92. The estimates from emissions-based DR will be significantly higher if bigger load curtailment and smaller load shifts are considered.

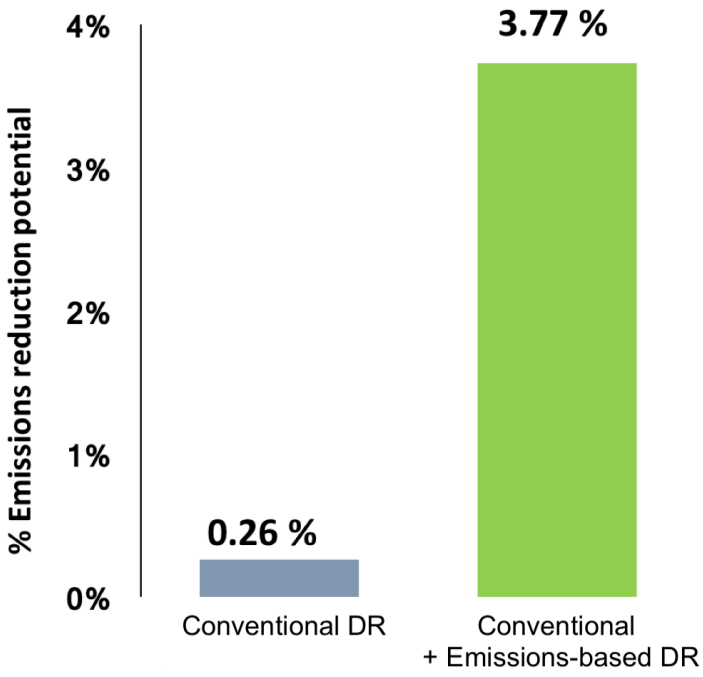

Figure 92: Emissions reduction potential through emissions-based DR layered on top of conventional DR

7.4 Conclusions

In this chapter, the use case for a reliable grid emissions data publicly available in real-time to tap into the emissions reduction potential of large consumer building flexible loads has been dealt with. The same methodology could be used to assess the emissions reduction potential of residential buildings.

Enabling the emissions-based DR for building consumers participating in conventional DR programs is easier since the existing communication and control systems could be effectively put in to use. Additionally, the data-driven pipelines facilitate collection of such data, their processing and availability to the consumers in the appropriate form.

Certain building consumers such as corporate companies are reluctant to participate in conventional DR programs, since the additional revenue through commitment of a large load curtailment may not be attractive to them. Nonetheless, the possibility of emissions reduction through the commitment of small load curtailment could help them achieve their corporate emission reduction targets. This could also work as a foot-in-the-door policy towards attracting such building consumers to participate in other DR programs.

To realise more participation in the emissions-based DR program, it is important to ensure that cost-effective smart devices are available at the consumer end. There is immense business opportunity for manufacturers of smart devices equipped with the required hardware and software to enable this. The role of system operators and utilities in helping materialise the emissions-based DR program is indispensable. In conclusion, concerted efforts are required to offer consumers the power to decide where their next unit of electricity is sourced from.

Chapter 8

Conclusions and further research

8.1 Thesis summary

The presented research demonstrated the application of data-driven models in demand response (DR) operational tasks such as capacity scheduling, reliable operation and performance evaluation for large consumer energy assets such as flexible loads, standby generators and storage systems. The research also showed that data-driven pipelines established through data collection, model development and deployment enable the use of new grid data such as real-time grid carbon emissions to develop an emissions-based DR program. This is expected to enable the consumers with flexible loads to achieve their emission reduction goals, a provision missing in the conventional DR programs.

Chapter-wise summaries are given below.

- Chapter 1 introduced the research presented in the thesis. The significance of data-driven modelling in DR, research objectives, scope and contributions were discussed.
- Chapter 2 provided a holistic perspective on DR and built a contextual platform for understanding the rest of the thesis. The scope of DR within smart grids, types of DR, DR program parameters, physical implementation of DR, and DR programs in different electricity markets have been examined.
- Chapter 3 surveyed the existing literature relevant to the research presented in the thesis. Building load estimation based on physics-based, hybrid and data-driven models were discussed. Machine learning (ML) concepts have been introduced with focus on supervised regression algorithms that were

implemented in data-driven building load estimation models. A comprehensive review of hyperparameter optimisation (HPO) methods used in data-driven building load estimation studies has been presented. Few of the data-driven modelling applications in the area of DR have been summarised. Battery energy storage systems (BESS) based DR applications were reviewed with state-of-charge (SoC) management strategies highlighted.

- Chapter 4 developed a computationally efficient, cost-effective and deployable data-driven model for building load estimation. Data pre-processing, model development and model deployment stages were addressed in great detail. An office building dataset has been used for demonstration purposes. Data pre-processing steps such as data collection, data cleaning, feature engineering, feature selection and data scaling were discussed. The selection of methods was justified and demonstrated. As part of the model development, a heuristic random-search HPO method was used to calibrate multiple ML algorithms such as artificial neural networks (ANN), support vector machines (SVM) and gradient boosted trees (GBT). Fine calibration was demonstrated through two stages of random search HPO method using small sampling sizes, thereby rendering the model computationally efficient. Data-driven modelling was demonstrated for day-ahead and week-ahead load estimation on the office building dataset and favourable predictive performances were achieved. The modelling approach developed in this chapter was then applied for DR operations in the succeeding two chapters.
- Chapter 5 demonstrated the application of the generic data-driven model for DR operational tasks such as: 1) *capacity scheduling* that facilitates better DR planning and 2) *performance evaluation* that measures and incentivises the actual consumer response during a DR event with reference to the scheduled capacity or baseline. Conventional models used for these DR tasks were observed to be non-standardised across the DR program operator regions. In this chapter, data-driven models using ML algorithms were developed for such DR tasks on multiple commercial buildings (supermarket, laboratory, hotel, retail store,

office, hospital) with load curtailment capability. These buildings were assumed to participate in ongoing incentive-based DR programs in the UK and the US electricity markets. The data-driven models were shown to perform better than the conventional models used in the respective markets.

- Chapter 6 demonstrated the application of data-driven models for SoC management of consumer BESS to ensure reliable operation in multiple DR programs. The SoC management strategy was developed for consumer BESS participation in a frequency responsive DR program (enhanced frequency response) and a peak shaving DR program (red-zone management) in the UK electricity market. A data-driven day ahead load estimation model was developed to forecast the consumer load for the peak shaving period. Using this model, the BESS SoC required prior to the peak shaving period (set-point SoC) is estimated one day in advance. This set-point SoC is achieved using the deadband allowance of the frequency responsive DR program. This data-driven SoC management strategy was simulated on a university building with Li-ion BESS. A sensitivity analysis was performed to test the robustness of this approach. An economic analysis was also performed to evaluate the feasibility of the project, which was shown to be viable. The assumptions used in the study and the simulation results were critically discussed.
- Chapter 7 showed the impact of introducing data-driven pipelines of data collection, model development and model deployment in the DR industry. The possibility of using new smart grid data to derive value out of it was discovered. A real-time grid marginal carbon emissions data available in public was used to design an emissions-based DR program; this is slowly being implemented in some of the electricity markets. The available flexibility in large consumer buildings already participating in conventional load curtailment DR programs was evaluated. A survey of large consumer building load flexibility from previous studies was also performed. Based on this, scenarios for participation in the emissions-based DR program were created for large consumer buildings in three different system operator regions in the US electricity market. It was

observed that, load curtailment per building up to 10% with 50% shift would still enable emission savings to the range of 3.77%. In aggregated scales, this translates to a large amount of avoided carbon from the grid.

8.2 Thesis statement validation

Thesis statement was: *"Data-driven models utilise smart grid data and enhance the demand response from large consumer energy assets such as flexible loads, standby generators and storage systems"*

In the work presented, a data-driven modelling approach was developed to utilise smart meter data for DR applications. The capability of data-driven models to improve DR related operational tasks such as capacity scheduling, reliable operation and performance evaluation was demonstrated based on large consumer energy assets and ongoing DR programs. Where applicable, the performances of data-driven models were compared with that of the conventional models and proven to be more reliable. The utilisation of real-time grid carbon emissions data in enabling an emissions-based DR program was discovered to be possible if data-driven pipelines are created. The results and inferences of the research hence validates the thesis statement.

8.3 Potential impact of the research

The impact of the research presented in this thesis are discussed in the sections below.

8.3.1 Effective use of the smart grid 'big data'

Large amounts of data are available in the smart grid. However, their potential for data-driven decision making is seldom realised in the electricity sector. The presented research demonstrates the use of large consumer smart meter data for developing data-driven models. It is also shown that data-driven pipelines hence

created, enable the use of new data such as real-time grid carbon emissions, that help building consumers reduce their emissions. The thesis builds confidence in the use of available data for different applications in the electricity sector, such as DR.

8.3.2 Solution to the new and impending challenges

DR programs have been in existence since the 1970's and the models used for operational tasks were based on classical approaches. The earlier versions of DR were focussed on consumer load curtailment for peak shaving only. The challenges of today's electricity systems are different. Especially with increased renewable penetration into the grid, and the impending electric vehicle deployment, there is a critical need for large volume, fast responding and predictable flexibility on the demand-side. Reliance on classical approaches towards modelling operational tasks may not be sufficient to meet such challenges, even though the DR industry is observed to favour such approaches. The thesis proposes a data-driven approach towards addressing these new and impending challenges. The demonstrated data-driven modelling also prioritises computational efficiency, allowing the models to be deployed in systems with low computational capabilities such as internet-of-things devices, ubiquitously.

8.3.3 Automation in the DR industry

The thesis demonstrates the capability of data-driven models in DR operational tasks such as capacity scheduling, reliable operation and performance evaluation. The data-driven model developed here is reliable, computationally efficient, cost-effective and replicable. This enables easy and fast development of data-driven models for a large number of consumers who participate in DR programs using their energy assets such as flexible loads, standby generators and storage systems. The use of data-driven modelling is a promising step towards automation in the DR industry. This is especially important in paradigms such as machine-to-machine (M2M) communication within smart grids that enables different devices such as

sensors, controllers and communication technologies to interact without human intervention.

8.3.4 Smart grid as a cyber-physical system (CPS)

A CPS is understood as a system that uses computational capability to interact with the physical environment supported by communication and control technologies. Smart grids are physical systems supported by communication and control technologies. In order to realise smart grid as a CPS, the available computational capability need to be used to interact with the physical system. Data-driven modelling collects data from the physical system (smart meter data), develops useful models using computational resources and communicates control actions to the physical system. Hence, data-driven modelling helps realise the potential of smart grids as a CPS, bringing in additional benefits to the electricity sector.

8.4 Dissemination

The thesis research and its findings were (or are being) disseminated to a wider audience through the following:

- ***Department for Business, Energy & Industrial Strategy (gov.uk), Innovative Non-Domestic Demand-Side Response Competition:*** Flexitricity has been selected for its project '*Quickturn'* which is a demonstration of low-cost and fast load curtailment DR in the UK electricity market. Data-driven modelling developed in this thesis is a key component of this project. As part of this, the computationally efficient building load estimation models are being deployed in DR operations. The innovative approach will also be disseminated to the public as mandated in the project guidelines.
- ***IEEE PES PowerTech conference, Manchester 2017 (presentation):*** This presentation titled *'Consumer – the emerging actor in grid balancing'* was part of a special session hosted by ADVANTAGE. The increasing importance of

consumer energy assets in electricity grid balancing and their flexibility potential was presented.

- ***Multi-stakeholder workshop, Chicago, 2017:*** The emissions reduction potential of commercial buildings based on grid emissions-based DR was estimated as part of a research exchange at the Rocky Mountain Institute in Boulder, Colorado, US. The results of this study were presented at a multi-stakeholder workshop in Chicago that was aimed at bringing together the representatives of the energy industry towards adoption of grid emissions-based DR in the US electricity markets.
- ***Research event at the Royal Society of London, 2016 (poster and presentation):*** The generic data-driven model for application in DR programs was communicated through a poster and presentation. The event enabled deeper interaction with smart grid enthusiasts and researchers.
- ***Various training events: Telecommunications Technology Centre of Catalonia (Spain, 2016), University of Novi Sad (Serbia, 2016), Aalborg University (Denmark, 2015):*** The progress of the research work from 2014 to 2016 was presented to the colleagues and partners involved in the ADVANTAGE project. Feedbacks were collected. Intellectually stimulating conversations with experts helped streamline the research work. Industry interactions enabled to stay up to date with the challenges in the electricity sector and their prospective solutions.

8.5 Limitations of the research and further work

There is enormous research interest in developing machine learning (ML) algorithms and data-driven modelling techniques around the world. The generic data-driven modelling developed in the thesis is demonstrated based on a selection of three ML algorithms. As part of further work, more ML algorithms are recommended to be used and calibrated on the building datasets, before model selection. Within the available computational capability, it is also recommended to widen the search space of trial hyperparameter values, to increase the sampling size

and to use more than two stages of random-search HPO for finer calibration of ML algorithms.

Sequential model-based methods [188–191] that in principle distinguish between important and unimportant dimensions during optimisation are being experimented in the empirical ML community, as an improvement over random-search method used in the generic data-driven model. The application of such HPO methods for development of finely calibrated data-driven models using the least computational resources and domain expertise requirement, need to be explored in the future.

The generic data-driven modelling is observed to have good performance on large consumer building loads with patterns influenced by temporal variables, weather variables and building variables. This was demonstrated largely on commercial consumers such as offices, retail stores, supermarket, laboratory, hotel, hospital and university buildings. Industrial consumer loads such as sewage treatment plants, manufacturing plants, etc. with inconsistent or random load patterns (as it appears during load profiling) are difficult to be learnt using supervised ML algorithms without knowing the cause of the variations. However, unsupervised ML algorithms may be used to extract information that is not available otherwise. This needs to be investigated in the future.

Residential building loads represent a significant portion of the overall electricity demand. In aggregated scales, load flexibility within residential buildings could be used for DR. The generic data-driven modelling approach presented in the thesis has not been evaluated for residential load estimation. It is expected that reliable, computationally efficient, cost-effective and replicable data-driven models as developed for large consumers here, may also help enhance DR from small consumers. Although markets such as the UK have not yet allowed DR participation for residential consumers, data-driven modelling could prove useful when the regulatory environment changes.

APPENDIX A:
DR programs in mainland Europe

Country – System operator	DR Program	Minimum capacity requirement	Notification	Duration	Number of events	Availability incentive	Utilisation incentive
Austria - APG	Primary control	2MW				22,01 €/MW/h	No
	Tertiary control			>4 hours			
Belgium - ELIA	Primary control: R1-Load (Up)	1MW	15s (50%), 30s (100%)		~80min/year	5-6 €/MW/h	No
	R3-Dynamic profile	1MW	15min	<2h	<40/year	3.07 €/MW/h	No
	Interruptiple clients	1MW	3min	4-12h	<4/year, 16-24 h/year	141 €/MW/h	75 €/MWh
	Strategic reserves		Day-ahead, 6.5h(warmup), 1.5h(rampdown)	1-12h	20-40/year or 130h/winter	Unknown	68 €/MWh
Denmark - Energienet	Primary Reserve	0.3MW	30s				
	Secondary Reserve	1MW	15mins				
	Frequency-controlled normal operation reserve	0.3MW	150s				
	Frequency-controlled disturbance reserve	0.3MW	5s (50%), 25s (100%)				
	Tertiary Reserve	10W	No				
	Strategic Reserves	200MW	>=10h				
Finland - Fingrid	Frequency-controlled normal operation reserve	0.1MW	3min			16.21 €/MW/h	No
	Frequency-controlled disturbance reserve	1MW/10MW	5s (50%), 30s(100%)		many/day , 1/year	4.13 €/MW/h, 0.5 €/MW/h	No
	Automatic frequency restoration reserve	5MW	2min		many/day	0	Yes
	Fast disturbance reserve - tertiary	10MW	15min		1/year	0.5 €/MW/h	580 €/MWh
	Strategic reserves	10MW	15min		1-2/winter	Unknown	
	Balancing Market - tertiary	10MW	15min			0	
	Wholesale market - Elspot,elbas	0.1MW			bids		
France - RTE	Primary control	1MW	<30s		many	160k€/MW/y	10.43 €/MWh
	Secondary control	1MW	<15min		many		
	Fast reserve	10MW	13min		many	36k€/MW/y	
	Complementary reserve	10MW	30min		many	21k€/MW/y	
	Demand response call for tender	10MW	2h		<60days/year	10-40k€/MW/y	
	Wholesale market-NEBEF	0.1MW	Day ahead		bids		
	Premium explicity DR ("l'effacement résidentiel diffus")						2-16 €/MWh
Germany	Primary control reserve	1MW	30s		many/day		
	Secondary control reserve	5MW	5min	12h	many/day		
	Minute reserve	5MW	15min	4h	many/day		
	Immediately	50MW	<1s			2500€/MW/mon	100-400€/MWh

	interruptiple loads					th	
	Quicly interruptible loads		<15min				
Ireland - Eirgrid	Short term active response (interruptiple load)	4MW	2s	5mins	10-20/year	No	8.20 €/MWh
	Demand side unit (price based capacity provision)	No	No		many/day	81.60 €/kW/year	No
Italy - Terna	Interruptible contract						
	Mainland fast	1MW	200ms		many	89899 €/MW	3000 €/MW
	Mainland emergency	1MW	5s		many	90000 €/MW	3000 €/MW
	Island fast	1MW	200ms		many	300000 €/MW	3000 €/MW
	Spot market	1MWh					
Netherland - Tennet	Primary reaction	1MW	<30s			Yes	No
	Regulating capacity	4MW	30s			Yes	€70/MWh
	Emergency power	20MW	<15min	<8h	40h/year or 19-27/year	Yes	€200/MWh
	Reserve capacity	4MW				No	€200/MWh
Norway	Frequency controlled normal operation reserve	5MW	50%(5s), 100%(30s)		many/hour	Yes	€42.03/MWh
	Frequency controlled disturbance reserve	5MW	50%(5s),100%(30s)		10000min/year	Yes	
	Automatic frequency restoration reserve	5MW	2min	<30min		Yes	
	Fast disturbance reserve	10MW	15min			€0.15-5/MWh	
	Balancing market	10MW	15min		many/day	No	€4-150/MWh
	Strategic reserves	10MW	4-48h		few/winter	No	No
	Energy options (interruptible loads)		7days				
Poland - PSE SA	Emergency demand response programme	10M	6h	2-4h	15/year or 2/day or 3/week	No	€220-276/MWh
	Balancing market	1MW				No	€30-50/MWh
Slovenia - ELES	Tertiary reserve	5MW	15min	<2h	35/year or 2/day	€38500/MW	€240/MWh
Spain - REE	Interruptible load: Mainland	5MW/90MW	Instant,15min or 2h		240-360h/y or 40-60h/month -	€97-176/MW	Yes
	Interruptible load: Islands	0.8MW	instant to 2h		120h/year		
Sweden - Svenska Kraftnat	Frequency containment reserves for normal operating band	0.1MW	63%(60s), 100%(3min)			Yes	Yes
	Frequency containment reserves for disturbances	1MW	50%(5s), 100%(30s)		500/year	Yes	Yes
	Automatic frequency restoration reserves	5MW	2min			Yes	Yes
	Fast disturbance reserve	5MW/10MW	15min			No	
	Strategic reserve / peak power reserve	5MW	15min		<10h/year	Yes	
	Balancing market	10MW	15min				
Switzerland - Swissgrid	Primary control	1MW	30s		many	23.14CHF/MW/h	No
	Secondary control	5MW	5min		many	28.28CHF/MW/h	
	Tertiary control	5MW	15min		many	3.84-5.64CHF/MW/h	

Bibliography

[1] Rousseaux F. BIG DATA and Data-Driven Intelligent Predictive Algorithms to support creativity in Industrial Engineering. Comput Ind Eng 2016;112:459–65. doi:10.1016/j.cie.2016.11.005.

[2] Ekanayake J, Liyanage K, Wu J, Yokoyama A, Jenkins N. Smart Grid: Technology and applications. Chichester, UK: John Wiley & Sons, Ltd; 2012. doi:10.1002/9781119968696.

[3] Zhou K, Fu C, Yang S. Big data driven smart energy management: From big data to big insights. Renew Sustain Energy Rev 2016;56:215–25. doi:10.1016/j.rser.2015.11.050.

[4] Goldenburg C, Dyson M, Masters H. Demand Flexibility: The key to enabling a low-cost, low-carbon grid 2018. https://www.rmi.org/wp-content/uploads/2018/02/Insight_Brief_Demand_Flexibility_2018.pdf (accessed February 10, 2018).

[5] Paterakis NG, Erdinç O, Catalão JPS. An overview of Demand Response: Key-elements and international experience. Renew Sustain Energy Rev 2017;69:871–91. doi:10.1016/j.rser.2016.11.167.

[6] Eia. World energy demand and economic outlook EIA's handling of non-U.S. policies in the International Energy Outlook. vol. 2016. 2016.

[7] Harish VSKV, Kumar A. A review on modeling and simulation of building energy systems. Renew Sustain Energy Rev 2016;56:1272–92. doi:10.1016/j.rser.2015.12.040.

[8] Kumar D, Singh H, Reshma. A review on industry challenges in smart grid implementation. 2016 7th India Int. Conf. Power Electron., IEEE; 2016, p. 1–5. doi:10.1109/IICPE.2016.8079395.

[9] Deng R, Yang Z, Chow M-Y, Chen J. A Survey on Demand Response in Smart Grids: Mathematical Models and Approaches. IEEE Trans Ind Informatics 2015;11:1–1. doi:10.1109/TII.2015.2414719.

[10] Vardakas JS, Zorba N, Verikoukis C V. A Survey on Demand Response Programs in Smart Grids: Pricing Methods and Optimization Algorithms. IEEE Commun Surv Tutorials 2015;17:152–78. doi:10.1109/COMST.2014.2341586.

[11] North American Electric Reliability Corporation (NERC). Data Collection for Demand-Side Management for Quantifying its Influence on Reliability. 2007.

[12] Federal Energy Regulatory Commission. Assessment of Demand Response and Advanced Metering February Staff Report 2011;February.

https://www.ferc.gov/industries/electric/indus-act/demand-response/dem-res-adv-metering.asp (accessed August 30, 2017).

[13] Siano P. Demand response and smart grids—A survey. Renew Sustain Energy Rev 2014;30:461–78. doi:10.1016/j.rser.2013.10.022.

[14] Eid C, Koliou E, Valles M, Reneses J, Hakvoort R. Time-based pricing and electricity demand response: Existing barriers and next steps. Util Policy 2016;40:15–25. doi:10.1016/j.jup.2016.04.001.

[15] Walawalkar R, Apt J, Mancini R. Economics of electric energy storage for energy arbitrage and regulation in New York. Energy Policy 2007;35:2558–68. doi:10.1016/j.enpol.2006.09.005.

[16] Grid N. System Operability Framework 2016 - Full non-interactive document 2016. http://www2.nationalgrid.com/UK/Industry-information/Future-of-Energy/System-Operability-Framework/ (accessed September 20, 2017).

[17] Dreidy M, Mokhlis H, Mekhilef S. Inertia response and frequency control techniques for renewable energy sources: A review. Renew Sustain Energy Rev 2017;69:144–55. doi:10.1016/j.rser.2016.11.170.

[18] Braithwait SD, Eakin K. The role of demand response in electric power market design. 2002.

[19] International Energy Agency. The power to choose - Demand response in liberalised electricity markets. 2003.

[20] Son J, Hara R, Kita H, Tanaka E. Operation scheduling considering demand response in a commercial building with chiller system and energy storage system. 2015 IEEE PES Asia-Pacific Power Energy Eng. Conf., IEEE; 2015, p. 1–5. doi:10.1109/APPEEC.2015.7381008.

[21] Deng R, Yang Z, Chow M-Y, Chen J. A Survey on Demand Response in Smart Grids: Mathematical Models and Approaches. IEEE Trans Ind Informatics 2015;11:570–82. doi:10.1109/TII.2015.2414719.

[22] National Grid. FFR testing guidance for providers 2017:1–21. http://www2.nationalgrid.com/uk/services/balancing-services/frequency-response/firm-frequency-response/ (accessed August 10, 2017).

[23] Electricity map - live CO2 emissions of electricity consumption n.d. https://www.electricitymap.org/?lang=en&wind=false&solar=false&page=map (accessed October 10, 2017).

[24] Hurley D, Peterson P, Whited M. Demand Response as a Power System Resource: Program Designs, Performance, and Lessons Learned in the United States. 2013.

[25] University of California Davis. Current energy - realtime grid status n.d. https://currentenergy.ucdavis.edu/ (accessed November 1, 2017).

[26] Energy F, Commission R. Assessment of demand response and advanced metering 2011. https://www.ferc.gov/industries/electric/indus-act/demand-response/dem-res-adv-metering.asp (accessed August 25, 2017).

[27] The Brattle Group, Freeman Sullivan & Co, Global Energy Partners. A National Assessment of Demand Response Potential. 2009. doi:10.1017/CBO9781107415324.004.

[28] EIA. Electric power sales, revenue, and energy efficiency Form EIA-861 detailed data files n.d. https://www.eia.gov/electricity/data/eia861/ (accessed January 15, 2017).

[29] National Grid. Non-BM Balancing Services and Volumes 2016:1–14. http://www2.nationalgrid.com/UK/Services/Balancing-services/Demand-Side-Response/ (accessed August 30, 2017).

[30] National Grid. Future Energy Scenarios 2016. http://fes.nationalgrid.com/ (accessed August 10, 2017).

[31] A smart, flexible energy system - A call for evidence - OFGEM report. Dep Business, Energy Ind Strateg 2016. https://www.ofgem.gov.uk/publications-and-updates/smart-flexible-energy-system-call-evidence (accessed August 25, 2017).

[32] Power responsive DSR product map glossary n.d. http://www2.nationalgrid.com/UK/Services/Balancing-services/Demand-Side-Response/ (accessed August 15, 2017).

[33] National Grid. Enhanced frequency response: Invitation to tender for pre-qualified parties 2016. http://www2.nationalgrid.com/Enhanced-Frequency-Response.aspx (accessed August 30, 2017).

[34] Smart Energy Demand Coalition (SEDC). Explicit Demand Response in Europe Mapping the Markets 2017 2017. http://www.smarten.eu/wp-content/uploads/2017/04/SEDC-Explicit-Demand-Response-in-Europe-Mapping-the-Markets-2017.pdf.

[35] Foucquier A, Robert S, Suard F, Stéphan L, Jay A. State of the art in building modelling and energy performances prediction: A review. Renew Sustain Energy Rev 2013;23:272–88. doi:10.1016/j.rser.2013.03.004.

[36] Coakley, Raftery P, Keane M. A review of methods to match building energy simulation models to measured data. Renew Sustain Energy Rev 2014;37:123–41. doi:10.1016/j.rser.2014.05.007.

[37] US Department of Energy Federal Energy Management Program. M & V Guidelines : Measurement and Verification for Contracts 2015. https://www.energy.gov/eere/femp/downloads/mv-guidelines-measurement-and-verification-performance-based-contracts-version.

[38] Koulamas C, Kalogeras AP, Pacheco-Torres R, Casillas J, Ferrarini L. Suitability analysis of modeling and assessment approaches in energy efficiency in buildings.

Energy Build 2018;158:1662–82. doi:10.1016/j.enbuild.2017.12.002.

[39] Saffari M, de Gracia A, Ushak S, Cabeza LF. Passive cooling of buildings with phase change materials using whole-building energy simulation tools: A review. Renew Sustain Energy Rev 2017;80:1239–55. doi:10.1016/j.rser.2017.05.139.

[40] Baloch AA, Shaikh PH, Shaikh F, Leghari ZH, Mirjat NH, Uqaili MA. Simulation tools application for artificial lighting in buildings. Renew Sustain Energy Rev 2018;82:3007–26. doi:http://dx.doi.org/10.1016/j.rser.2017.10.035.

[41] Yang XS, Zhao LH, Bruse M, Meng QL. An integrated simulation method for building energy performance assessment in urban environments. Energy Build 2012;54:243–51. doi:DOI 10.1016/j.enbuild.2012.07.042.

[42] Yildiz B, Bilbao JI, Sproul AB. A review and analysis of regression and machine learning models on commercial building electricity load forecasting 2017;73:1104–22. doi:10.1016/j.rser.2017.02.023.

[43] Mat Daut MA, Hassan MY, Abdullah H, Rahman HA, Abdullah MP, Hussin F. Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. Renew Sustain Energy Rev 2017;70:1108–18. doi:10.1016/j.rser.2016.12.015.

[44] Deb C, Zhang F, Yang J, Eang Lee S, Wei Shah K. A review on time series forecasting techniques for building energy consumption 2017;74:902–24. doi:10.1016/j.rser.2017.02.085.

[45] Ghahramani Z. Unsupervised Learning. Adv. Lect. Mach. Learn., vol. 3176, 2004, p. 72–112. doi:10.1007/978-3-540-28650-9_5.

[46] Pedregosa F, Grisel O, Weiss R, Passos A, Brucher M. Scikit-learn : Machine Learning in Python. J Mach Learn Res 2011;12:2825–30. doi:10.1007/s13398-014-0173-7.2.

[47] Amasyali K, Gohary N. A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev 2018;81:1192–205. doi:10.1016/J.RSER.2017.04.095.

[48] Amazon. Amazon Machine Learning Developer Guide 2015. https://docs.aws.amazon.com/machine-learning/latest/dg/machinelearning-dg.pdf#importance-of-feature-transformation (accessed March 1, 2018).

[49] Developers S-L. Scikit-Learn User Guide Release 0.18.2 2017. http://scikit-learn.org/0.18/_downloads/scikit-learn-docs.pdf (accessed February 5, 2018).

[50] Kuhn M, Johnson K. Applied Predictive Modeling. Springer US; 2013. doi:10.1007/978-1-4614-6849-3.

[51] Xypolytou E, Meisel M, Sauter T. Short-term electricity consumption forecast with artificial neural networks — A case study of office buildings. 2017 IEEE Manchester PowerTech, IEEE; 2017, p. 1–6. doi:10.1109/PTC.2017.7980874.

[52] Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. Ann Stat 2001;29:1189–232.

[53] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. ArXiv 2016:1–6. doi:10.1145/2939672.2939785.

[54] Microsoft. Azure machine learning n.d. https://docs.microsoft.com/en-us/azure/machine-learning/studio/algorithm-choice (accessed November 1, 2017).

[55] Rasmussen CE, Williams CKI, Processes G, Press MIT, Jordan MI. Gaussian Processes for Machine Learning. 2006.

[56] Mining D. Springer Series in Statistics The Elements of. Math Intell 2009;27:83–85. doi:10.1007/b94608.

[57] Gibson A, Patterson J. Deep learning. O'Reilly Media, Inc.; 2017.

[58] Wang Z, Srinivasan RS. A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. Renew Sustain Energy Rev 2016;2016–Febru:3438–48. doi:10.1016/j.rser.2016.10.079.

[59] Raza MQ, Khosravi A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renew Sustain Energy Rev 2015;50:1352–72. doi:10.1016/j.rser.2015.04.065.

[60] Dong B, Cao C, Lee SE. Applying support vector machines to predict building energy consumption in tropical region. Energy Build 2005;37:545–53. doi:10.1016/j.enbuild.2004.09.009.

[61] Heo Y, Zavala VM. Gaussian process modeling for measurement and verification of building energy savings. Energy Build 2012;53:7–18. doi:10.1016/j.enbuild.2012.06.024.

[62] Burkhart MC, Heo Y, Zavala VM. Measurement and verification of building systems under uncertain data: A Gaussian process modeling approach. Energy Build 2014;75:189–98. doi:10.1016/j.enbuild.2014.01.048.

[63] Peng Y, Rysanek A, Nagy Z, Schlüter A. Using machine learning techniques for occupancy-prediction-based cooling control in o ffi ce buildings. Appl Energy 2018;211:1343–58. doi:10.1016/j.apenergy.2017.12.002.

[64] Drgoňa J, Picard D, Kvasnica M, Helsen L. Approximate model predictive building control via machine learning. Appl Energy 2018;218:199–216. doi:10.1016/j.apenergy.2018.02.156.

[65] Guo Y, Wang J, Chen H, Li G, Liu J, Xu C, et al. Machine learning-based thermal response time ahead energy demand prediction for building heating systems. Appl Energy 2018;221:16–27. doi:10.1016/j.apenergy.2018.03.125.

[66] Chae YT, Horesh R, Hwang Y, Lee YM. Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. Energy Build

2016;111:184–94. doi:10.1016/j.enbuild.2015.11.045.

[67] Jetcheva JG, Majidpour M, Chen WP. Neural network model ensembles for building-level electricity load forecasts. Energy Build 2014;84:214–23. doi:10.1016/j.enbuild.2014.08.004.

[68] Leung MC, Tse NCF, Lai LL, Chow TT. The use of occupancy space electrical power demand in building cooling load prediction. Energy Build 2012;55:151–63. doi:10.1016/j.enbuild.2012.08.032.

[69] Lam JC, Wan KKW, Wong SL, Lam TNT. Principal component analysis and long-term building energy simulation correlation. Energy Convers Manag 2010;51:135–9. doi:10.1016/j.enconman.2009.09.004.

[70] Ahmad AS, Hassan MY, Abdullah MP, Rahman HA, Hussin F, Abdullah H, et al. A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renew Sustain Energy Rev 2014;33:102–9. doi:10.1016/j.rser.2014.01.069.

[71] Zhao H, Magoulès F. A review on the prediction of building energy consumption. Renew Sustain Energy Rev 2012;16:3586–92. doi:10.1016/j.rser.2012.02.049.

[72] Shcherbakov M, Kamaev V. Automated Electric Energy Consumption Forecasting System Based On Decision Tree Approach. vol. 46. IFAC; 2013. doi:10.3182/20130619-3-RU-3018.00486.

[73] Yu Z, Haghighat F, Fung BCM, Yoshino H. A decision tree method for building energy demand modeling. Energy Build 2010;42:1637–46. doi:10.1016/j.enbuild.2010.04.006.

[74] Tso GKF, Yau KKW. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy 2007;32:1761–8. doi:10.1016/j.energy.2006.11.010.

[75] Kolter JZ, Ferreira J. A Large-scale Study on Predicting and Contextualizing Building Energy Usage. Proc Conf Artif Intell (AAAI), Spec Track Comput Sustain AI, 2011 2011:8.

[76] Noh HY, Rajagopal R. Data-driven forecasting algorithms for building energy consumption. Proc SPIE - Int Soc Opt Eng 2013;8692:86920T. doi:10.1117/12.2009894.

[77] Valgaev O. Building Power Demand Forecasting Using K-Nearest Neighbors Model - Initial Approach. IEEE PES Asia-Pacific Power Energy Conf., Xi'an, China: 2016, p. 1055–60.

[78] Naji S, Keivani A, Shamshirband S, Alengaram UJ, Jumaat MZ, Mansor Z, et al. Estimating building energy consumption using extreme learning machine method. Energy 2016;97:506–16. doi:10.1016/j.energy.2015.11.037.

[79] Li S, Goel L, Wang P. An ensemble approach for short-term load forecasting by extreme learning machine. Appl Energy 2016;170:22–9.

doi:10.1016/j.apenergy.2016.02.114.

[80] Fan C, Xiao F, Zhao Y. A short-term building cooling load prediction method using deep learning algorithms 2017;195:222–33. doi:10.1016/j.apenergy.2017.03.064.

[81] Bergstra J, Bengio Y. Random Search for Hyper-Parameter Optimization. J Mach Learn Res 2012;13:281–305.

[82] Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for Hyper-Parameter Optimization. Adv Neural Inf Process Syst 2011:2546–54. doi:2012arXiv1206.2944S.

[83] Chae YT, Horesh R, Hwang Y, Lee YM. Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. Energy Build 2016;111:184–94. doi:10.1016/j.enbuild.2015.11.045.

[84] Chou JS, Bui DK. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. Energy Build 2014;82:437–46. doi:10.1016/j.enbuild.2014.07.036.

[85] Sörensen K. Metaheuristics-the metaphor exposed. Int Trans Oper Res 2015;22:3–18. doi:10.1111/itor.12001.

[86] Ruiz L, Cuéllar M, Calvo-Flores M, Jiménez M. An Application of Non-Linear Autoregressive Neural Networks to Predict Energy Consumption in Public Buildings. Energies 2016;9:684. doi:10.3390/en9090684.

[87] Li Q, Meng Q, Cai J, Yoshino H, Mochida A. Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks. Energy Convers Manag 2009;50:90–6. doi:10.1016/j.enconman.2008.08.033.

[88] Chen B-J, Chang M-WM-W, Lin C-JC-J. Load Forecasting Using Support Vector Machines: A Study on EUNITE Competition 2001. IEEE Trans Power Syst 2004;19:1821–30. doi:10.1109/TPWRS.2004.835679.

[89] Shu F, Luonan C. Short-term load forecasting based on an adaptive hybrid method. Power Syst IEEE Trans 2006;21:392–401. doi:10.1109/TPWRS.2005.860944.

[90] Larochelle H, Erhan D, Courville A, Bergstra J, Bengio Y. An empirical evaluation of deep architectures on problems with many factors of variation. Proc. 24th Int. Conf. Mach. Learn. - ICML '07, New York, New York, USA: ACM Press; 2007, p. 473–80. doi:10.1145/1273496.1273556.

[91] Massana J, Pous C, Burgas L, Melendez J, Colomer J. Short-term load forecasting in a non-residential building contrasting models and attributes. Energy Build 2015;92:322–30. doi:10.1016/j.enbuild.2015.02.007.

[92] Jain RK, Smith KM, Culligan PJ, Taylor JE. Forecasting energy consumption of multi-family residential buildings using support vector regression : Investigating the impact of temporal and spatial monitoring granularity on per ... Forecasting energy consumption of multi-family residential buildings. Appl Energy 2014;123:168–78.

doi:10.1016/j.apenergy.2014.02.057.

[93] Kontokosta CE, Tull C. A data-driven predictive model of city-scale energy use in buildings 2017;197:303–17. doi:https://doi.org/10.1016/j.apenergy.2017.04.005.

[94] Massana J, Pous C, Burgas L, Melendez J, Colomer J. Short-term load forecasting for non-residential buildings contrasting artificial occupancy attributes. Energy Build 2016;130:519–31. doi:10.1016/j.enbuild.2016.08.081.

[95] Liu D, Chen Q. Prediction of building lighting energy consumption based on support vector regression. 2013 9th Asian Control Conf ASCC 2013 2013. doi:10.1109/ASCC.2013.6606376.

[96] Zhao D, Zhong M, Zhang X, Su X. Energy consumption predicting model of VRV (Variable refrigerant volume) system in office buildings based on data mining. Energy 2016;102:660–8. doi:10.1016/j.energy.2016.02.134.

[97] Ghasemi A, Shayeghi H, Moradzadeh M, Nooshyar M. A novel hybrid algorithm for electricity price and load forecasting in smart grids with demand-side management. Appl Energy 2016;177:40–59. doi:10.1016/j.apenergy.2016.05.083.

[98] Liu N, Tang Q, Zhang J, Fan W, Liu J. A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids. Appl Energy 2014;129:336–45. doi:10.1016/j.apenergy.2014.05.023.

[99] AlRashidi MR, EL-Naggar KM. Long term electric load forecasting based on particle swarm optimization. Appl Energy 2010;87:320–6. doi:10.1016/j.apenergy.2009.04.024.

[100] Chou JS, Ngo NT. Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns. Appl Energy 2016;177:751–70. doi:10.1016/j.apenergy.2016.05.074.

[101] Afshin M, Sadeghiant A, Raahemifar K. On efficient tuning of LS-SVM hyper-parameters in short-term load forecasting: A comparative study. 2007 IEEE Power Eng Soc Gen Meet PES 2007. doi:10.1109/PES.2007.385613.

[102] Nghiem TX, Jones CN. Data-driven demand response modeling and control of buildings with Gaussian Processes. 2017 Am Control Conf 2017:2919–24. doi:10.23919/ACC.2017.7963394.

[103] Jung D, Krishna VB, Temple WG, Yau DKY. Data-driven evaluation of building demand response capacity. 2014 IEEE Int Conf Smart Grid Commun SmartGridComm 2014 2015:541–7. doi:10.1109/SmartGridComm.2014.7007703.

[104] Behl M, Smarra F, Mangharam R. DR-Advisor: A data-driven demand response recommender system. Appl Energy 2016;170:30–46. doi:10.1016/j.apenergy.2016.02.090.

[105] Chen Y, Xu P, Chu Y, Li W, Wu Y, Ni L, et al. Short-term electrical load forecasting

using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings. Appl Energy 2017;195:659–70. doi:10.1016/j.apenergy.2017.03.034.

[106] Park S, Ryu S, Choi Y, Kim J, Kim H. Data-driven baseline estimation of residential buildings for demand response. Energies 2015;8:10239–59. doi:10.3390/en80910239.

[107] The Energyst. Battery Storage Report 2017 2017. https://theenergyst.com/storage/ (accessed October 11, 2017).

[108] Oudalov A, Chartouni D, Ohler C. Optimizing a Battery Energy Storage System for Primary Frequency Control. IEEE Trans Power Syst 2007;22:1259–66. doi:10.1109/TPWRS.2007.901459.

[109] Thorbergsson E, Knap V, Swierczynski M, Stroe D, Teodorescu R. Primary Frequency Regulation with Li-Ion Battery Based Energy Storage System - Evaluation and Comparison of Different Control Strategies. Telecommun Energy Conf 'Smart Power Effic (INTELEC), Proc 2013 35th Int 2013:1–6.

[110] Borsche T, Ulbig A, Koller M, Andersson G. Power and energy capacity requirements of storages providing frequency control reserves. IEEE Power Energy Soc. Gen. Meet., Vancouver: 2013. doi:10.1109/PESMG.2013.6672843.

[111] Borsche TS, Ulbig A, Andersson G. Impact of frequency control reserve provision by storage systems on power system operation. IFAC Proc., vol. 19, 2014, p. 4038–43.

[112] D. Tretheway. Regulation Energy Management Draft Final Proposal. CAISO 2011. https://www.caiso.com/Documents/RevisedDraftFinalProposal-RegulationEnergyManagement-Jan13_2011.pdf (accessed September 15, 2017).

[113] Lian B, Sims A, Yu D, Wang C, Dunn RW. Optimizing LiFePO4 Battery Energy Storage Systems for Frequency Response in the UK System. IEEE Trans Sustain Energy 2017;8:385–94. doi:10.1109/TSTE.2016.2600274.

[114] Megel O, Mathieu JL, Andersson G. Maximizing the potential of energy storage to provide fast frequency control. IEEE PES ISGT Eur. 2013, IEEE; 2013, p. 1–5. doi:10.1109/ISGTEurope.2013.6695380.

[115] Zhai Q, Ma J, Dong Z, Lu C. Dynamic restoration of the state of charge for the battery participating in mutiple markets using market price signals. 2015 5th Int. Conf. Electr. Util. Deregul. Restruct. Power Technol., IEEE; 2015, p. 103–7. doi:10.1109/DRPT.2015.7432212.

[116] Mariaud A, Acha S, Ekins-Daukes N, Shah N. Integrated Optimisation of PV and Battery Storage Systems for UK Non- Domestic buildings. ASHRAE Annu Conf Proceedings, June 24 - June 28 2017;199:1–8.

[117] Greenwood D, Wade N, Taylor P, Papadopoulos P, Heyward N, Alilat S. A Forecasting, Optimization and Scheduling System for Energy Storage Systems in distribution networks. 2016 IEEE Power Energy Soc. Gen. Meet., IEEE; 2016, p. 1–5.

doi:10.1109/PESGM.2016.7741684.

[118] Ulbig A, Galus MD, Chatzivasileiadis S, Andersson G. General frequency control with aggregated control reserve capacity from time-varying sources: The case of PHEVs. 2010 IREP Symp - Bulk Power Syst Dyn Control - VIII, IREP2010 2010. doi:10.1109/IREP.2010.5563261.

[119] Ulbig A, Rinke T, Chatzivasileiadis S, Andersson G. Predictive control for real-time frequency regulation and rotational inertia provision in power systems. Proc IEEE Conf Decis Control 2013:2946–53. doi:10.1109/CDC.2013.6760331.

[120] Mishra S, Sahoo S, Pullaguram DR. A Systematic State of Charge based V2G Charging Framework for Frequency Response. IFAC-PapersOnLine, vol. 48, 2015, p. 31–6. doi:10.1016/j.ifacol.2015.12.349.

[121] Jovanovic R, Sretenovic A. Various multistage ensembles for prediction of heating energy consumption. Model Identif Control A Nor Res Bull 2015;36:119–32. doi:10.4173/mic.2015.2.4.

[122] Jovanović R, Sretenović AA, Živković BD. Ensemble of various neural networks for prediction of heating energy consumption. Energy Build 2015;94:189–99. doi:10.1016/j.enbuild.2015.02.052.

[123] Drucker H. Improving regressors using boosting techniques. 14th Int Conf Mach Learn 1997:107–15.

[124] Hassan MA, Khalil A, Kaseb S, Kassem MA. Exploring the potential of tree-based ensemble methods in solar radiation modeling. Appl Energy 2017;203:897–916. doi:10.1016/j.apenergy.2017.06.104.

[125] Bansal A, Rompikuntla SK, Gopinadhan J, Kaur A, Kazi ZA. Energy Consumption Forecasting for Smart Meters. arXiv1512.05979v1 BAI Conf. 2015, Bangalore: 2015.

[126] Hunn BD. Fundamentals of building energy dynamics. Sol Heat Technol 1996;4:538. http://books.google.de/books?id=puZl5Wk1ICkC.

[127] Arens EA, Williams PB. The effect of wind on energy consumption in buildings. Energy Build 1977;1. doi:10.1016/0378-7788(77)90014-7.

[128] Hodge VJ, Austin J. A Survey of Outlier Detection Methodologies. Artif Intell Rev 2004;22:85–126. doi:10.1007/s10462-004-4304-y.

[129] Ben-gal I. Outlier detection. In: Maimon O, Rokach L, editors. Data Min. Knowl. Discov. Handb., Boston, MA: Springer US; 2010, p. 1–16. doi:10.1007/978-0-387-09823-4.

[130] John W. Tukey. Exploratory data analysis 1977:530–7. https://www.stat.berkeley.edu/~brill/Papers/EDASage.pdf (accessed February 10, 2018).

[131] Moya MM, Hush DR. Network constraints and multi-objective optimization for one-

class classification. Neural Networks 1996;9:463–74. doi:10.1016/0893-6080(95)00120-4.

[132] Dreiseitl S, Osl M, Scheibböck C, Binder M. Outlier Detection with One-Class SVMs: An Application to Melanoma Prognosis. AMIA Annu Symp Proc 2010;2010:172–6.

[133] Lyu S, Farid H. Steganalysis using color wavelet statistics and one-class support vector machines. SPIE Symp Electron Imaging 2004;5306:35–46. doi:10.1117/12.526012.

[134] Manevitz LM, Yousef M, Cristianini N, Shawe-Taylor J, Williamson B. One-Class SVMs for Document Classification. J Mach Learn Res 2001;2:139–54. doi:10.1162/15324430260185574.

[135] Senf A, Chen X, Zhang A. Comparison of One-Class SVM and Two-Class SVM for Fold Recognition. Neural Inf Process 2006:140–9. doi:10.1007/11893257_16.

[136] Orloff J, Bloom J. Bootstrap confidence intervals. Introd to Probab Stat - MIT OpenCourseWare 2014. https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading24.pdf (accessed April 10, 2018).

[137] Mourshed M. Relationship between annual mean temperature and degree-days. Energy Build 2012;54:418–25. doi:10.1016/j.enbuild.2012.07.024.

[138] Met Office. Global accuracy at a local level 2018. https://www.metoffice.gov.uk/about-us/who/accuracy/forecasts (accessed May 25, 2018).

[139] Ofgem. Review of Met Office weather forecast accuracy 2018. https://www.ofgem.gov.uk/ofgem-publications/111122 (accessed May 20, 2018).

[140] Minitab. Weather Forecasts : Just how reliable are they ? 2015:1–7. http://www.minitab.com/uploadedFiles/Content/News/Published_Articles/WeatherForecasts-HowReliable-EN.pdf (accessed May 20, 2018).

[141] Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. Bioinformatics 2007;23:2507–17. doi:10.1093/bioinformatics/btm344.

[142] Shong N. Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data. Winona state university, 2010. doi:10.1017/CBO9781107415324.004.

[143] Vergara JR, Estévez PA. A review of feature selection methods based on mutual information. Neural Comput Appl 2014;24:175–86. doi:10.1007/s00521-013-1368-0.

[144] Shcherbakov MV, Brebels A, Shcherbakova NL, Tyukov AP, Janovsky TA, Kamaev VA evich. A survey of forecast error measures. World Appl Sci J 2013;24:171–6. doi:10.5829/idosi.wasj.2013.24.itmies.80032.

[145] Fausett L V. Fundamentals of Neural Networks: Architectures, Algorithms and Applications. 1st ed. Pearson; 1993.

[146] Biswas MAR, Robinson MD, Fumo N. Prediction of residential building energy consumption: A neural network approach. Energy 2016;117:84–92. doi:10.1016/j.energy.2016.10.066.

[147] Magoulas GD, Vrahatis MN. Adaptive algorithms for neural network supervised learning: a deterministic optimization approach. Int J Bifurc Chaos 2006;16:1929–50. doi:10.1142/S0218127406015805.

[148] Amami R, Ben Ayed D, Ellouze N. Practical Selection of SVM Supervised Parameters with Different Feature Representations for Vowel Recognition. Int J Digit Content Technol Its Appl 2015;7:418–24. doi:10.4156/jdcta.vol7.issue9.50.

[149] Li Q, Meng Q. Development and application of hourly building cooling load prediction model. 2010 Int Conf Adv Energy Eng 2010:392–5. doi:10.1109/ICAEE.2010.5557536.

[150] Loh W-Y. Classification and regression trees. Wiley Interdiscip Rev Data Min Knowl Discov 2011;1:14–23. doi:10.1002/widm.8.

[151] Li P, Wang H, Zhang B. A Distributed Online Pricing Strategy for Demand Response Programs. IEEE Trans Smart Grid 2017:1–1. doi:10.1109/TSG.2017.2739021.

[152] Mohajeryami S, Doostan M, Asadinejad A, Schwarz P. Error analysis of customer baseline load (CBL) calculation methods for residential customers. IEEE Trans Ind Appl 2017;53:5–14. doi:10.1109/TIA.2016.2613985.

[153] Bode J, Sullivan M, Eto JH. Measuring Short-term Air Conditioner Demand Reductions for Operations and Settlement. 2012.

[154] Grid N. System Needs and Product Strategy 2017. http://www2.nationalgrid.com/UK/Services/Balancing-services/Future-of-balancing-services/ (accessed August 25, 2017).

[155] Grid N. Frequency Control by Demand Management Version 1.1 n.d. http://www2.nationalgrid.com/WorkArea/DownloadAsset.aspx?id=41294 (accessed August 15, 2017).

[156] Menke R, Abraham E, Parpas P, Stoianov I. Demonstrating demand response from water distribution system through pump scheduling. Appl Energy 2016;170:377–87. doi:10.1016/j.apenergy.2016.02.136.

[157] ERCOT. Technical requirements and scope of work (Emergency Response Service) 2017. http://www.ercot.com/services/programs/load/eils (accessed August 10, 2017).

[158] Parianen S, Oree V. Investigating the use of standby generator sets to mitigate peak demand charges in industrial facilities. 2015 IEEE 15th Int. Conf. Environ. Electr. Eng., IEEE; 2015, p. 659–64. doi:10.1109/EEEIC.2015.7165243.

[159] Cottuli C. Comparison of Static and Rotary UPS - Schneider Electric white paper n.d. http://www.apc.com/salestools/DBOY-78KRZE/DBOY-78KRZE_R2_EN.pdf (accessed

August 25, 2017).

[160] Macdonald J, Kiliccote S, Boch J, Chen J, Nawy R. Commercial Building Loads Providing Ancillary Services in PJM. 2014.

[161] Kwon PS, Østergaard P. Assessment and evaluation of flexible demand in a Danish future energy scenario. Appl Energy 2014;134:309–20. doi:10.1016/j.apenergy.2014.08.044.

[162] Piette MA, Watson DS, Motegi N, Bourassa N. Findings from the 2004 fully automated demand response tests in large facilities. 2005.

[163] Kiliccote S, Lanzisera S, Liao A, Schetrit O, Piette MA. Fast DR : Controlling Small Loads over the Internet. 2014.

[164] Yin R, Kara EC, Li Y, DeForest N, Wang K, Yong T, et al. Quantifying flexibility of commercial and residential loads for demand response using setpoint changes. Appl Energy 2016;177:149–64. doi:10.1016/j.apenergy.2016.05.090.

[165] Hu J, Cao J, Yong T, Guerrero JM, Chen MZQ, Li Y. Demand Response Load Following of Source and Load Systems. IEEE Trans Control Syst Technol 2017;25:1586–98. doi:10.1109/TCST.2016.2615087.

[166] Mathieu JL, Price PN, Kiliccote S, Piette MA. Quantifying changes in building electricity use, with application to demand response. IEEE Trans Smart Grid 2011;2:507–18. doi:10.1109/TSG.2011.2145010.

[167] ERCOT. Default baseline methodology 2016. http://www.ercot.com/services/programs/load/eils/documents (accessed August 10, 2017).

[168] Vartanian C. Grid stability battery systems for renewable energy success. 2010 IEEE Energy Convers Congr Expo ECCE 2010 - Proc 2010:132–5. doi:10.1109/ECCE.2010.5618059.

[169] Wade NS, Taylor PC, Lang PD, Jones PR. Evaluating the benefits of an electrical energy storage system in a future smart grid. Energy Policy 2010;38:7180–8. doi:10.1016/j.enpol.2010.07.045.

[170] Koch M, Krüger M, Tenbohlen S. Frequency Response from Autonomous Battery Energy Storage. CIGRE US Natl. Comm. 2012 Grid Futur. Symp., 2012, p. 1–6. doi:10.1109/PESGM.2012.6345361.

[171] Koller M, Borsche T, Ulbig A, Andersson G. Review of grid applications with the Zurich 1 MW battery energy storage system. Electr Power Syst Res 2015;120:128–35. doi:10.1016/j.epsr.2014.06.023.

[172] Madawala UK, Thrimawithana DJ. A Bidirectional Inductive Power Interface for Electric Vehicles in V2G Systems. IEEE Trans Ind Electron 2011;58:4789–96. doi:10.1109/TIE.2011.2114312.

[173] Mu Y, Wu J, Ekanayake J, Jenkins N, Jia H. Primary frequency response from electric vehicles in the Great Britain power system. IEEE Trans Smart Grid 2013;4:1142–50. doi:10.1109/TSG.2012.2220867.

[174] Wu D, Kintner-Meyer M, Yang T, Balducci P. Analytical sizing methods for behind-the-meter battery storage. J Energy Storage 2017;12:297–304. doi:10.1016/j.est.2017.04.009.

[175] Plc SEPN. Use of System Charging Statement 2017. http://www.ukpowernetworks.co.uk/internet/en/about-us/duos/ (accessed September 10, 2017).

[176] Battery discharge cycles. Apple Inc n.d. https://www.apple.com/batteries/why-lithium-ion/ (accessed August 30, 2017).

[177] Valence battery n.d. https://www.valence.com/why-valence/long-lifecycle/ (accessed September 20, 2017).

[178] Lariviere J, Baker HH, Mccormick G, Kawano S, Cheek JG, Johnson JE, et al. How Better Accounting Can More Cheaply Reduce Carbon Emissions Baker Center Staff 2016.

[179] Carter TH, Wang C, Miller SS, McElmurry SP, Miller CJ, Hutt IA. Modeling of power generation pollutant emissions based on locational marginal prices for sustainable water delivery. IEEE 2011 EnergyTech, ENERGYTECH 2011 2011:1–6. doi:10.1109/EnergyTech.2011.5948499.

[180] Rogers MM, Wang Y, Wang C, McElmurry SP, Miller CJ. Evaluation of a rapid LMP-based approach for calculating marginal unit emissions. Appl Energy 2013;111:812–20. doi:10.1016/j.apenergy.2013.05.057.

[181] Watttime. WattTime grid emissions data n.d. https://api.watttime.org/ (accessed March 15, 2017).

[182] Federal Energy Regulatory Comission. Electric power markets n.d. https://www.ferc.gov/market-oversight/mkt-electric/overview.asp (accessed March 24, 2017).

[183] Kiliccote S, Piette MA, Mathieu J, Parrish K. Findings from Seven Years of Field Performance Data for Automated Demand Response in Commercial Buildings. 2010.

[184] Agnew K, Burke R, Ham-su P. Participation of Demand Response Resources in ISO New England's Ancillary Service Markets. Energy Progr Eval Conf 2007:220–33.

[185] Massachussets Electric. Retail mall DR - case study. Natl Grid 2004:2004. https://www.nationalgridus.com/media/pdfs/billing-payments/shared_demand_response_mall_casestudy.pdf (accessed October 30, 2017).

[186] Broad D, Christensen K, Dragoon K, Lee K, Spielman K, Parkinson S. Smart End-use

Energy Storage and Integration of Renewable Energy A Pilot Project Overview. ECOFYS - Sustain Energy Everyone 2012.

[187] Ghatikar G, Ganti V, Matson N. Demand Response Opportunities and Enabling Technologies for Data Centers : Findings from Field Studies. Lawrence Berkeley Natl Lab 2012.

[188] Bardenet R, Brendel M, Kégl B, Sebag M. Collaborative hyperparameter tuning. Proc 30th Int Conf Mach Learn 2013;28:199–207.

[189] Martinez-Cantin R. BayesOpt: A Bayesian optimization library for nonlinear optimization, experimental design and bandits. J Mach Learn Res 2014;15:3735–9.

[190] Bergstra J, Yamins DLK, Cox DD. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. Proc 30th Int Conf Mach Learn 2013:115–23.

[191] Hutter F, Hoos HH, Leyton-Brown K. Sequential Model-Based Optimization for General Algorithm Configuration 2011:507–23. doi:10.1007/978-3-642-25566-3_40.